2025 International Conference On Rehabilitation Robotics (ICORR)
May 12-16, 2025. Chicago, IL, USA

Optimizing Cross-Modal Matching for Multimodal Motor Rehabilitation

Anway S. Pimpalkar
Dept. of Biomedical Engineering
Johns Hopkins University
Baltimore, MD, USA
apimpal1@jhu.edu

A. Michael West Jr.
Dept. of Mechanical Engineering
Johns Hopkins University
Baltimore, MD, USA
awest36@jhu.edu

Jing Xu
Dept. of Kinesiology
University of Georgia
Athens, GA, USA
jing.xu@uga.edu

Jeremy D. Brown
Dept. of Mechanical Engineering
Johns Hopkins University
Baltimore, MD, USA
jdelainebrown@jhu.edu

2025 International Conference On Rehabilitation Robotics (ICORR) | 979-8-3503-8068-2/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICORR66766.2025.11063112

***Abstract*—Stroke often causes sensorimotor deficits, impairing hand dexterity and disrupting independence for millions worldwide. While rehabilitation devices leveraging visual and haptic feedback show promise, their effectiveness is limited by a lack of perceptual equity, which is necessary to ensure fair comparisons between sensory modalities. This study refines cross-modal matching protocols to address this gap, enabling unbiased evaluation of multimodal feedback. Using the Hand Articulation and Neurotraining Device (HAND), 12 healthy participants matched visual and haptic stimuli in a structured task. A streamlined protocol, requiring just 2–3 blocks and 3 reference intensities, reduced experimental time fivefold while preserving data integrity. Data were analyzed using linear and exponential models applied to both full and reduced datasets. The results demonstrated consistent participant performance across trials, with higher matching errors at greater stimulus intensities, likely attributable to sensory saturation effects. Furthermore, the study offered practical methodological insights, including the use of reduced data sampling paradigms to enhance experimental efficiency significantly while preserving data integrity. This work advances perceptual equity in multisensory feedback systems, addressing sensory encoding variability to support scalable, personalized therapeutic strategies for stroke recovery.**

***Index Terms*—Cross-modal Matching, Perception, Haptic Feedback, Vibrotactile Feedback, Dexterity, Rehabilitation**

I. Introduction

The human hand is fundamental to performing essential activities of daily living, such as grasping, manipulating objects, and executing precise movements. However, hand dexterity is often compromised by neurological conditions like stroke, which impacts over 12.2 million individuals globally each year [1]. The loss of hand functionality not only reduces independence and quality of life but also places significant physical, emotional, and economic burdens on individuals [2].

Efforts to address this challenge have driven the development of innovative strategies to restore hand dexterity post-stroke. Among these approaches, hand rehabilitation devices have shown significant promise by leveraging advancements in robotics, neurorehabilitation, and haptic feedback to promote motor recovery and neural plasticity [3], [4]. Despite these advances, challenges remain in ensuring the accessibility, personalization, and effectiveness of these devices to meet the diverse needs of stroke patients [5].

The Hand Articulation and Neurotraining Device (HAND) was co-developed by Carducci, Xu, and their collaborators [6], [7] to explore individuated finger control and multidimensional co-activation patterns following stroke. Here, we focus on adapting the HAND to study the pinch grasp, a critical fine motor skill required for handling delicate or brittle objects. Pinch grip deficits, often observed in stroke patients, are linked to impairments in sensorimotor processing pathways. These deficits result in diminished tactile and proprioceptive feedback, reducing the ability to accurately modulate grip forces and perform everyday tasks independently [8].

While many existing rehabilitation devices rely heavily on visual feedback for training and recovery [9], evidence suggests that visual feedback alone is insufficient to optimize motor recovery outcomes [10]. Recent studies highlight the potential of integrating multisensory feedback, particularly haptic cues, to promote neural plasticity and enhance functional motor rehabilitation [11]. In this work, we aim to understand how multisensory feedback should best be provided in the HAND by utilizing visual and haptic cues to augment somatosensory pathways. By addressing these gaps, this study contributes to advancing therapeutic strategies and optimizing neurorehabilitation outcomes for stroke patients.

Sensory and motor deficits arising from stroke manifest in diverse and highly individualized ways, resulting in variations in how stimuli are perceived across individuals [10]. To effectively leverage multisensory feedback in rehabilitation, it is crucial for sensorimotor tasks to account for such variability and ensure fairness across modalities. Establishing *perceptual equity* – ensuring that any observed differences in performance between visual and haptic feedback reflect true differences in sensory encoding rather than biases introduced by the experimental setup — is essential for accurately comparing the efficacy of these modalities in healthy participants. This process, known as cross-modal matching, involves calibrating sensory intensities to establish a balance between modalities, enabling fair and unbiased comparisons [12], [13].

Cross-modal matching provides a standardized framework to evaluate the encoding and integration of sensory information across modalities, ensuring perceptual equity by aligning stim-

This work was supported by the National Institute of Neurological Disorders and Stroke (NINDS) of the NIH under Award Number R01NS130210 and by the Johns Hopkins University Provost's Postdoctoral Fellowship.

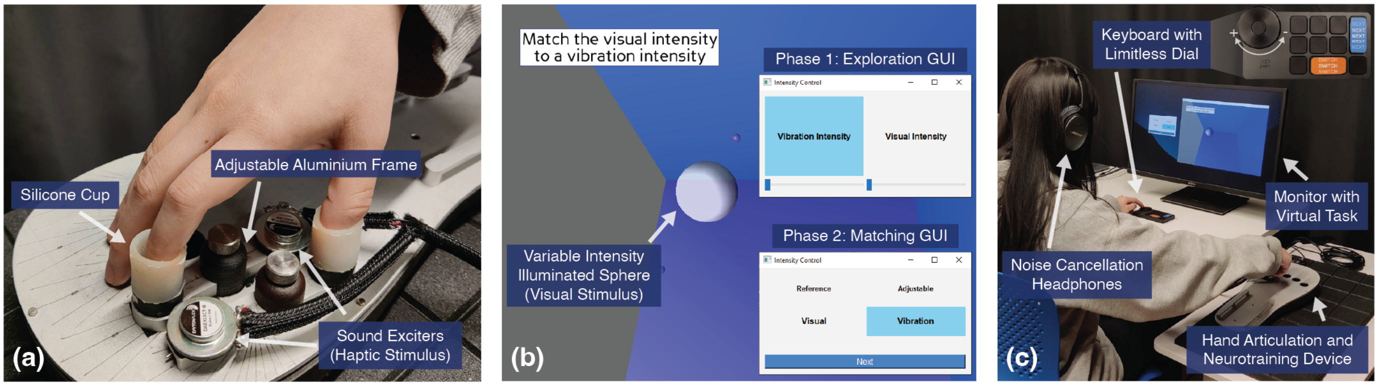

Fig. 1. A modified Hand Articulation and Neurotraining Device (HAND) [6], [7] setup used in this study. (a) Participants interact with the device by placing their fingers in silicone cups mounted on an aluminum frame, with vibrating voice coils providing mechanical haptic stimuli. (b) The virtual task includes a variable-brightness sphere for visual feedback and two user interfaces: the exploration phase, where reference sliders are present, and the matching phase, where no reference sliders are displayed. (c) Wearing noise-canceling headphones, participants use a limitless keyboard dial to adjust the intensity of the presented feedback. Feedback is either haptic – delivered through the HAND or visual – rendered within a virtual task.

ulus intensities. This process is crucial for isolating genuine differences in sensory encoding from biases introduced by experimental setups. However, it can be time-consuming and variable across individuals, which poses a significant challenge, particularly for stroke patients who often experience fatigue during extended tasks [14]. To address these issues, one of the critical aims of this study is to develop pre-experiment cross-modal matching protocols that streamline the process, minimize fatigue, and maintain accuracy. We believe the guidelines derived from this study with healthy individuals will provide a foundational framework for developing efficient, reliable methods in neurorehabilitation research with stroke patients, seamlessly integrating with existing clinical techniques.

II. Methods and Implementation

A. Participants

We evaluated the ability of N=12 individuals with no history of stroke or upper-extremity disabilities (5 males and 7 females; mean age: 25.2±6.1 years; 9 right-handed, 1 left-handed, and 2 ambidextrous) to match the intensity of one feedback modality to an equivalent intensity of another using a custom rehabilitation interface under two distinct feedback conditions. Each participant's handedness was assessed using the Laterality Quotient derived from the administered Edinburgh Handedness Inventory after obtaining their consent [15]. The experiment lasted approximately 60 minutes, and participants were compensated at a rate of $10 per hour. All participants provided informed consent in accordance with a protocol approved by the Johns Hopkins School of Medicine Institutional Review Board (Study #IRB00209185).

B. Experimental Setup

The study utilized a modified Hand Articulation and Neurotraining Device (HAND) built and validated in prior studies [6], [7]. The HAND includes two silicone cups mounted on an aluminum frame, precisely positioned to naturally align with each participant's index finger and thumb, allowing for a customized and comfortable fit as shown in Fig. 1(a). The device is set up for the participants' right hand, and all participants, regardless of handedness, use the same configuration. Each cup is mechanically coupled with a Dayton Audio DAEX13CT-8 Sound Voice Coil, which provides the vibrotactile haptic feedback necessary for the tasks.

The voice coils are driven through a custom amplifier board controlled by a 600 MHz Teensy 4.0 microcontroller. This microcontroller communicates with the primary control software, a Python script running on the computer via a serial interface. The virtual task program transmits 9-bit values to the Teensy, which adjusts the intensity of the voice coil output. Upon receiving these values, the microcontroller employs a MAX521BCPP digital-to-analog converter (Maxim Integrated) and a TPA3122D2 Class-D audio power amplifier (Texas Instruments) to generate 250 Hz sine waves with a variable peak-to-peak voltage range of 0 to 7 V. These signals drive the voice coils, delivering haptic stimuli to participants.

The virtual task consists of a 3D environment featuring a centrally positioned, fixed-size gray sphere illuminated by a distant light source, as shown in Fig. 1(b). The sphere's brightness is adjusted to provide visual stimuli, matching the 9-bit resolution used for the haptic stimuli generated by the voice coils. The virtual environment is created in Python 3.6 and rendered using Panda3D Open Source Framework.

C. Cross-modal Matching Implementation

To ensure fairness and perceptual equity in neurorehabilitation experiments, cross-modal matching is essential for calibrating stimulus intensities across modalities, ensuring that observed differences reflect genuine sensory encoding rather than experimental bias. Building on this principle, we implemented a cross-modal matching task based on the paradigm developed by Pitts et al. [12], [13], constructing participant-specific psychometric curves to align haptic and visual stimulus intensities. The participants were instructed to sit in front of a monitor running the virtual task and rest their right hand comfortably on the HAND as shown in Fig.1(c). Participants wore noise-canceling headphones that played 60 Hz white noise at a comfortable volume to eliminate potential auditory interference from the voice coils. The experiment was broken into two phases.

1) Intensity Exploration: In the first phase, participants used a graphical user interface (GUI) to explore the full range of vibrotactile and visual stimulus intensities for two minutes. The stimuli were presented continuously, with their intensities adjustable via a limitless dial on an XPPen ACK05 wireless shortcut keyboard (XPPen, China) as shown in Fig. 1(c). Two sliders on the screen indicated stimulus intensity being presented in real time. A clockwise click on the dial increased the intensity, while a counterclockwise click decreased it by 1 unit within a range of 0 to 255. A designated button on the keyboard allowed participants to toggle between the two modalities during the exploration phase, enabling them to choose which modality to adjust. Participants were instructed to familiarize themselves with the full ranges of the haptic and visual stimuli and their correlation, using the sliders as a reference. They were informed that the sliders would not be available in the subsequent phase.

2) Intensity Matching: In the intensity matching phase, participants were tasked with aligning the intensity of one sensory modality (haptic or visual) to a reference intensity provided in the other modality. This involved two distinct conditions: a haptic reference modality, where participants matched the visual stimulus to a given haptic intensity, and a visual reference modality, where they matched the haptic stimulus to a given visual intensity.

Following this setup, participants completed two experimental blocks in each condition. Specifically, participants used the same limitless dial configuration as in the intensity exploration phase to adjust the intensity of one stimulus to match a given reference stimulus. However, in this phase, the sliders were no longer displayed. Participants were allotted 90 seconds to submit their matched intensity for each trial, after which the current value was automatically recorded if the time limit was exceeded. Eight reference intensities, normalized and evenly distributed within the range of 0 to 1 (excluding endpoints), were used. The 8 reference intensities were shown 4 times in a block randomized order for a total of 32 trials. The sequence of reference intensities was maintained across both blocks. The order in which haptic and visual reference stimuli were presented, whether the haptic or visual modality served as the reference first, was counterbalanced across participants.

Upon completing each block, participants completed the NASA Task Load Index (NASA-TLX) questionnaire to evaluate cognitive workload [16]. At the conclusion of the session, participants were asked for a description of their strategies.

D. Data Analysis

We obtained two datasets for each participant: (1) the visual reference experiment and (2) the haptic reference experiment. Before analysis, the matched intensity was normalized such that values ranged from 0 to 1. This enabled a one-to-one relationship between the two modalities. The following analyses of these datasets will aim to (1) compare participants' performance across the two experimental conditions and (2) develop a protocol to help future researchers find participant-specific models that ensure perceptual equity in multimodal motor control experiments.

1) Comparing Dataset and Fit: We compared two different modeling approaches for deriving participant-specific models: a linear model and an exponential model. The linear model was selected because the experimental protocol inherently established a linear relationship between the haptic and visual modalities. Additionally, we included the exponential model because it has been employed in prior research [13], [17]. Thus, for each participant and experiment, the observed data were fit to both linear and exponential models using MATLAB 2024 b's *fitlm* and *fitnlm* functions, respectively.

Prior to model fitting, we applied two distinct data sampling paradigms. In the first paradigm, referred to as "All Data," we used all 32 data points collected across the four blocks, each consisting of eight reference intensities. In the second paradigm, referred to as "Average + Origin," we averaged the data across blocks at the eight reference intensities and added a ninth data point at the origin, as it is assumed that participants can identify when no reference intensity is provided. This second approach aligns with the methodology used in [17].

In summary, for each participant and each experiment, we generated a total of four models, combining two datasets ("All Data" and "Average + Origin") with two fits (linear and exponential). For each model, we report the coefficient of determination, R^2, as a measure of the goodness of fit. To compare each model's ability to sufficiently describe the data, we run a 2×2 (2 datasets × 2 fits) ANOVA on R^2.

2) Finding the Optimal Reduced Fit: To optimize the cross-modal matching procedure for efficiency, we aimed to determine the minimum number of blocks and reference intensities needed to achieve a model fit comparable to that obtained with the full experimental dataset. To this end, we conducted a Monte Carlo simulation for each participant, experiment, and fit. This simulation fit the data under 657 different conditions, derived as follows: we considered three block configurations (2, 3, or 4 blocks, preserving their sequential order to simulate truncated experiments) and all possible subsets of 3 to 8 reference intensities for each stimulus. The subsets of reference intensities were determined using the binomial coefficient $\binom{n}{k}$, where $n = 8$ and k ranged from 3 to 8. Note that a minimum of three reference points was required for fitting a linear model. This resulted in $\sum_{k=3}^{8}\binom{8}{k} = 219$ combinations of reference intensities. Multiplying these 219 combinations by the three block configurations yielded 657 total conditions. By systematically varying these parameters, we aimed to identify an experimental design that balances efficiency and accuracy by minimizing the cost function:

$$cost = \frac{blocks}{4} \cdot \frac{intensities}{8} \cdot 10 \cdot (x_1\ error + x_2\ error) \quad (1)$$

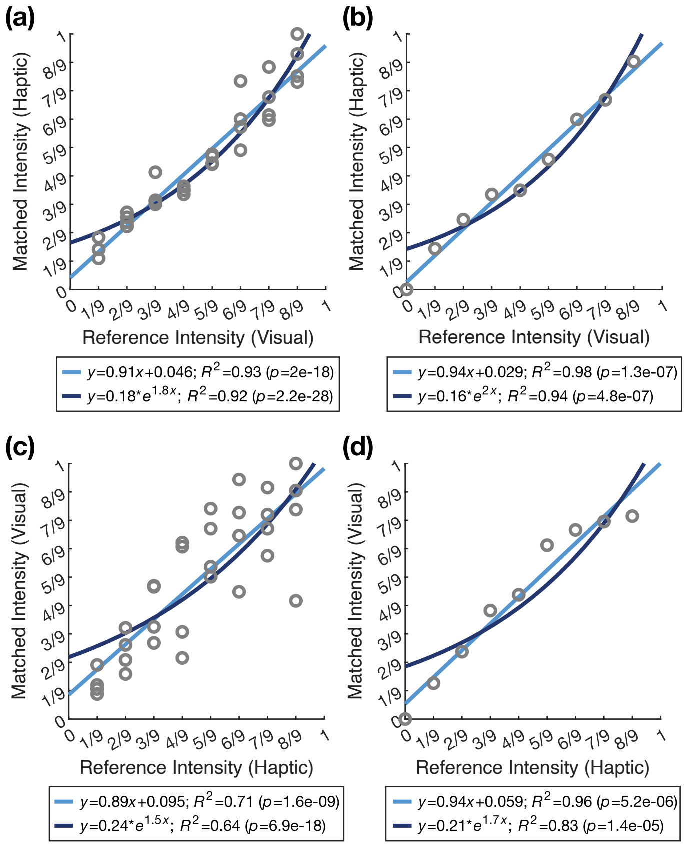

Fig. 2. Linear and exponential fits for a representative participant (Participant 10) in the cross-modal matching experiment. Subplots include: (a) visual modality as the reference (All Data), (b) visual modality as the reference (Average + Origin), (c) haptic modality as the reference (All Data), and (d) haptic modality as the reference (Average + Origin).

where $blocks$ is the number of blocks used, $intensities$ is the number of intensities used. x_1 $error$ and x_2 $error$ represent the deviation between the model's coefficients fitted to the subset of data and the model's coefficients fitted to the full dataset. A weighting factor of 10 was applied to the error term to heavily penalize deviations from the original fit heavily, ensuring the reduced dataset is consistent with the full dataset's fit.

After running the simulation, we identified the best fit for each case as the one that minimized the cost function defined in Eq. 1. This process yielded 24 optimal fits (12 participants $\times$ 2 experiments) for each dataset. To compare the characteristics of these optimally reduced fits, we conducted a paired Student's t-test to evaluate differences in the number of intensities, number of blocks, x_1 error, x_2 error, cost, and R^2 across the datasets.

3) Task Performance: We sought to understand how participants' performance varied between the two experiments. To quantify performance, we calculated error as the absolute difference between the normalized matched intensity and the normalized reference intensity $error = |matched - reference|$. Recall that in Section II-C, the relationship between the two

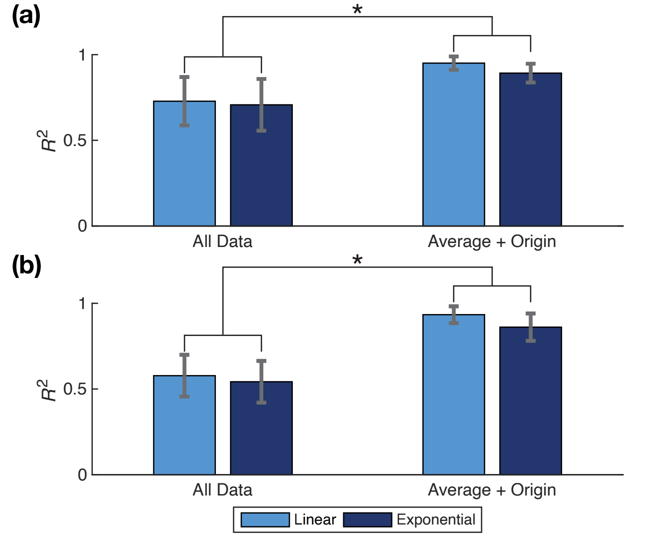

Fig. 3. Average coefficient of determination, $R^2, \pm$ 1 standard deviation for each dataset and fit across participants, with (a) visual modality as the reference and (b) haptic modality as the reference.

modalities was inherently linear; after normalization, it became one-to-one. If participants understood the underlying relationship perfectly, their normalized matched intensity would equal the normalized reference intensity.

To compare performance across experiments, we conducted a paired Student's t-test on the average error for each participant (averaged across all trials within each experiment). Additionally, we tested whether reference intensity or fatigue influenced performance. For reference intensity, we averaged each participant's error across trials at each intensity level and fit a linear model to error as a function of reference intensity. A significant slope would indicate that reference intensity affects performance. For fatigue, we averaged each participant's error across trials within each block and fit a linear model to error as a function of the block. A significant positive slope would indicate a decline in performance over time.

We analyzed NASA-TLX data using a two-way ANOVA to evaluate the effects of reference modality and category (mental demand, physical demand, temporal demand, effort, frustration, and perceived performance) on subjective workload ratings. If a significant effect was found, post-hoc paired Students' t-test was used to identify any significant differences between subcategories.

All data processing and statistical analyses were performed using custom scripts in MATLAB 2024. The significance level for statistical tests was $\alpha = 0.05$.

III. Results

A. Comparing Dataset and Fit

Fig. 2 illustrates the linear and exponential fits for a representative participant, providing a visual comparison of model performance. The models were assessed using the coefficient of determination, R^2, as a measure of goodness of fit.

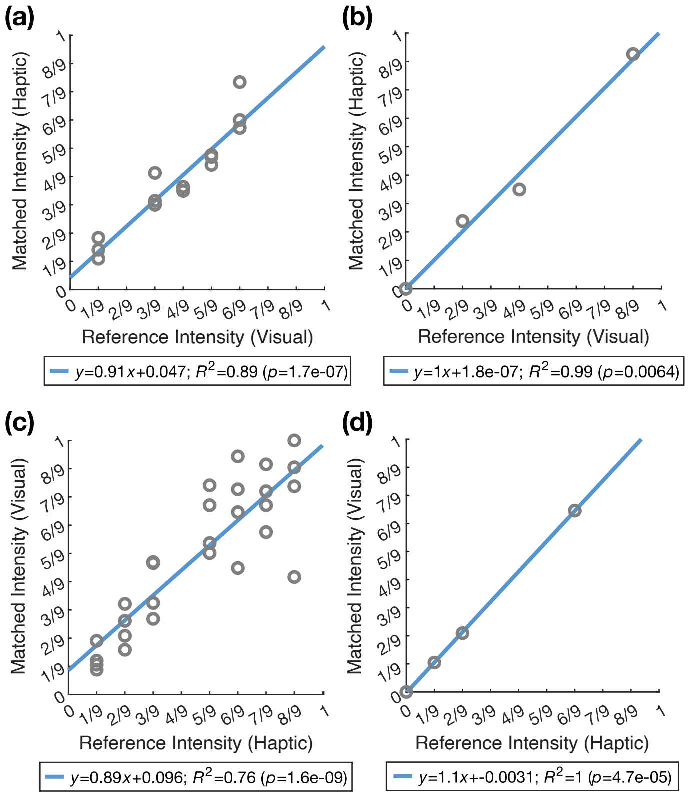

Fig. 4. Optimal reduced fit example for a representative participant (Participant 10), highlighting the reduced blocks and reference intensities. Subplots include: (a) visual modality as the reference (All Data), (b) visual modality as the reference (Average + Origin), (c) haptic modality as the reference (All Data), and (d) haptic modality as the reference (Average + Origin).

Fig. 3 shows the average R^2 across participants separated by experiment, dataset, and fit. Table I shows how each of the 4 models performed, in describing a participant's experimental data. Of note, all models were statistically significant. A 2×2 (2 datasets × 2 fits) ANOVA revealed a significant effect of dataset ($F_{1,95} = 6.458, p = 0.0127$); there was not a significant effect of fit ($p > 0.05$) nor a significant interaction ($p > 0.05$). However, the linear fit generally outperformed the exponential fit.

B. Finding the Optimal Reduced Fit

While the previous analysis included two datasets and two fits, we focused here on the linear fit for succinctness, as it generally performed better than the exponential fit and aligned with the underlying linear relationship in the task. An example for participant 10 is shown in Fig. 4.

Across the 24 optimal reduced fits produced for each of the Linear "All Data" and linear "Average + Origin" data we compared the number of intensities, number of blocks, x_1 error, x_2 error, cost and R^2 across the datasets. A paired Student's t-test revealed a significant effect of dataset on number of blocks ("All": M=3.54, SD=0.88; "Average + Origin": M=2.38, SD=1.06; p<0.001), number of intensities ("All": M=5.42, SD=1.44; "Average + Origin": M=3, SD=0; p<0.001), x_1 error ("All": M=1.4e-03, SD=1.6e-03; "Average + Origin": M=5.2e-03, SD=5.1e-03; p=1.19e-03), x_2 error ("All": M=1.2e-03, SD=1.7e-03; "Average + Origin": M=4.8e-03, SD=7.3e-03; p=2.19e-02), and R^2 ("All": M=0.66, SD=0.17; "Average + Origin": M=0.97, SD=0.022; p<0.001). However, there was not a significant effect on cost ("All": M=0.10, SD=8.9e-02; "Average + Origin": M=0.15, SD=0.11; p=0.10).

TABLE I

THE COEFFICIENT OF DETERMINATION, R^2, OF EACH MODEL FIT TO EACH PARTICIPANT'S DATA SEPARATED BY EXPERIMENT, DATASET, AND FIT. THE ABBREVIATIONS 'ALL,' AND 'AVG + O,' DENOTE THE ALL DATA AND AVERAGE + ORIGIN DATASETS, RESPECTIVELY. THE ABBREVIATIONS 'LIN,' AND 'EXP,' DENOTE THE LINEAR AND EXPONENTIAL FITS, RESPECTIVELY. AVG AND SD DENOTE THE AVERAGE AND STANDARD DEVIATIONS ACROSS PARTICIPANTS.

	Visual Reference				Haptic Reference			
Data Type	**All**		**Avg + O**		**All**		**Avg + O**	
Fit Type	**Lin**	**Exp**	**Lin**	**Exp**	**Lin**	**Exp**	**Lin**	**Exp**
P 1	0.51	0.47	0.93	0.82	0.56	0.54	0.97	0.95
P 2	0.78	0.81	0.96	0.97	0.57	0.58	0.97	0.95
P 3	0.74	0.69	0.94	0.88	0.52	0.44	0.80	0.68
P 4	0.89	0.87	0.99	0.92	0.53	0.51	0.92	0.89
P 5	0.89	0.87	0.99	0.94	0.69	0.65	0.98	0.93
P 6	0.82	0.76	0.98	0.90	0.45	0.39	0.94	0.82
P 7	0.62	0.57	0.95	0.86	0.30	0.29	0.91	0.84
P 8	0.65	0.65	0.98	0.93	0.55	0.49	0.90	0.78
P 9	0.52	0.50	0.87	0.78	0.71	0.70	0.94	0.93
P 10	0.93	0.92	0.98	0.94	0.71	0.64	0.96	0.83
P 11	0.76	0.79	0.90	0.92	0.64	0.63	0.97	0.91
P 12	0.63	0.60	0.92	0.85	0.70	0.63	0.94	0.83
Avg	0.73	0.71	0.95	0.89	0.58	0.54	0.93	0.86
SD	0.14	0.14	0.04	0.05	0.12	0.12	0.05	0.08

TABLE II

PERCENTAGE OF TIMES EACH REFERENCE INTENSITY APPEARED IN THE 24 OPTIMAL REDUCED FITS. 'ALL,' AND 'AVG + O,' DENOTE THE ALL DATA AND AVERAGE + ORIGIN DATASETS.

	Reference Intensity							
	1/9	**2/9**	**3/9**	**4/9**	**5/9**	**6/9**	**7/9**	**8/9**
All	0.71	0.58	0.63	0.58	0.75	0.67	0.75	0.75
Avg + O	0.58	0.29	0.33	0.42	0.42	0.38	0.25	0.33

Table II shows the percentage of times each reference intensity appeared in the 24 optimal reduced fits. In the "All Data" dataset, edge reference intensities (e.g., 1/9 and 8/9) were the most frequently used, but overall, the dataset exhibited some uniformity by often including more than five intensities. In contrast, the "Average + Origin" dataset used an average of approximately 3.54 reference intensities, and mostly included the intensities 1/9, 4/9, 5/9, and 6/9, in that order. This dataset generally excluded higher reference intensities.

C. Performance

Participants' performance was quantified by calculating the error as the absolute difference between the normalized matched intensity and the normalized reference intensity. Then, a paired Student's t-test was conducted to compare average errors across the two experiments for each participant. The test revealed no significant effect of the experiment on performance ($p > 0.05$). However, participants performed better in the visual reference experiment, as seen in Fig. 5.

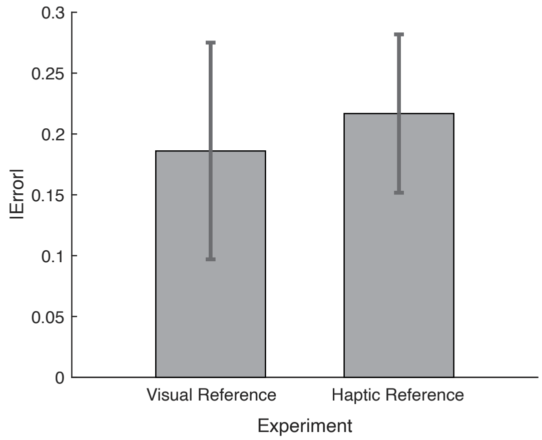

Fig. 5. Absolute error comparison between visual and haptic reference conditions.

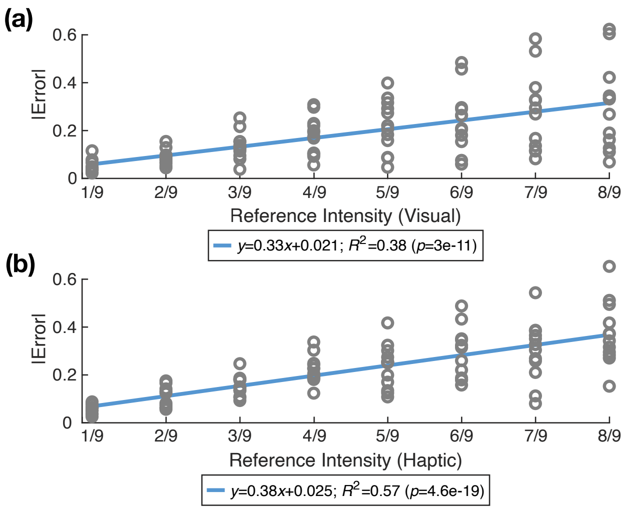

Fig. 6. Participant error as a function of reference intensity, demonstrating increased error at higher intensities. Subplots include: (a) visual modality as the reference and (b) haptic modality as the reference.

To examine the impact of reference intensity on performance, a linear model was fit to the error as a function of reference intensity. In both experiments, the model returned a significant positive slope (visual reference: M=0.33, $p<0.001$ and haptic reference: M=0.38, p=<0.001) indicating that participants performed worse at higher reference intensities, as shown in Fig. 6.

Similarly, to evaluate the impact of fatigue, a linear model was fit to the error as a function of block. In both experiments, the model did not yield a significant slope (visual reference: M=5.7e-04, $p>0.05$ and haptic reference: M=1.4e-06, $p>0.05$), indicating that participants' performance neither decreased nor increased as the experiment progressed, as shown in Fig. 7. This suggests that participants did not suffer from fatigue over the course of the study.

Finally, the two-way ANOVA on workload ratings revealed no statistically significant differences in NASA-TLX scores

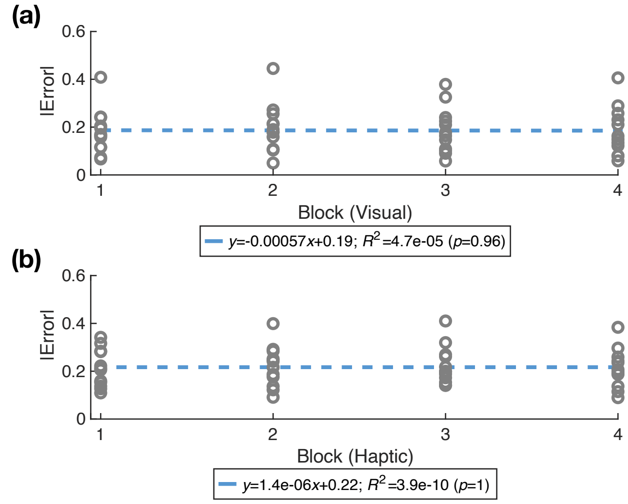

Fig. 7. Participant error as a function of block, showing stable performance throughout the experiment. Subplots include: (a) visual modality as the reference and (b) haptic modality as the reference.

TABLE III
NASA TLX SCORES FOR EACH WORKLOAD DIMENSION ACROSS THE VISUAL AND HAPTIC REFERENCE CONDITIONS. AVG AND SD DENOTE THE AVERAGE AND STANDARD DEVIATIONS ACROSS PARTICIPANTS.

	Visual Reference	**Haptic Reference**
Mental Demand	-1.25 ± 3.72	0.75 ± 4.77
Physical Demand	-3.33 ± 3.72	-2.58 ± 5.37
Temporal Demand	-5.75 ± 4.18	-6.00 ± 4.35
Effort	0.17 ± 4.06	2.83 ± 3.66
Frustration	-4.50 ± 5.16	-3.08 ± 3.66
Perceived Performance	11.58 ± 5.16	9.58 ± 2.94

across task categories or reference modalities. Additionally, there was no significant interaction. The survey data is summarized in Table III.

IV. Discussion

This study refined cross-modal matching procedures to enhance perceptual equity in multimodal psychophysics and non-psychophysics experiments by systematically evaluating and optimizing experimental parameters. Twelve participants completed two matching experiments using the HAND [6], [7], aligning visual and haptic stimuli according to a visual or haptic reference (Fig. 1). Data were analyzed using two datasets (All Data" and Average + Origin") and two fits (linear and exponential). The linear fit consistently outperformed the exponential fit, aligning with the task's inherent linear relationship. Simulations identified the minimum number of blocks and reference intensities needed to achieve comparable model fits to the full dataset. The All Data" approach utilized a broader range of intensities, while Average + Origin" favored low-to-mid-range intensities, excluding higher ones. Participants showed greater error at higher intensities but exhibited no signs of fatigue throughout the experiment. These findings provide actionable guidelines for streamlining cross-modal matching protocols while maintaining accuracy and efficiency, contributing to developing effective rehabilitation protocols.

When comparing goodness of fit (R^2), we found that the 'Average + Origin' dataset statistically outperformed the 'All Data' dataset (Table I). This result is supported by mathematical reasoning. The 'Average + Origin' dataset uses fewer data points and averages across blocks, resulting in one matched (output) point for each reference (input) point (Fig 2). This reduction in variance around the fit line increases R^2, producing an artificial advantage. It underscores a trade-off: improved model performance comes at the cost of ignoring trial-level variability. Importantly, the 'Average + Origin' dataset captures the overall trend between modalities, the primary goal of cross-modal matching. By prioritizing the overarching relationship, this approach ensures perceptual equity without being hindered by trial-level noise.

While the 'Average + Origin' method modeled the data well for healthy participants, it assumes no perceptual bias at zero intensity. This assumption may not hold in clinical populations, such as stroke survivors, who often experience sensory deficits or altered perceptual baselines. Testing this paradigm with individuals exhibiting perceptual biases could reveal whether the 'All Data' method provides better fits. Such findings could refine the protocol and offer deeper insights into how sensory encoding variability impacts calibration methods across populations. These considerations will guide future studies to improve the adaptability and robustness of cross-modal matching protocols for motor rehabilitation.

The optimal reduced order fits generated through the Monte Carlo simulation revealed an important observation: while the "Average + Origin" dataset achieved a significantly worse x_1 and x_2 error compared to the "All Data" dataset, the overall cost was not significantly different between the two datasets (refer to Section III-B. Upon closer inspection, the errors in both datasets were quite small (x_1 error: $\sim 1.4 \times 10^{-3}$ and $\sim 5.2 \times 10^{-3}$ for "All Data" and "Average + Origin," respectively; x_2 error: $\sim 1.2 \times 10^{-3}$ and $\sim 4.8 \times 10^{-3}$, respectively). This indicates that the 'Average + Origin' dataset, while slightly sacrificing fit accuracy compared to the non-reduced dataset, achieved this with fewer blocks and reference intensities. The reduced order fit suggested 2-3 blocks and 3 reference intensities. While participants were allotted 1 minute per trial, they completed each trial in an average of 12 seconds. Based on this performance, a full experiment following the reduced order fit, including both reference modalities, would take less than 5 minutes. Even if participants used the full allotted time, the reduced protocol would take only 12 to 18 minutes, significantly shorter than the original maximum length of 64 minutes (4 blocks × 8 reference intensities × 2 experiments). The result is a more efficient protocol that still captures the cross-modal relationship between modalities. This trade-off highlights the utility of the 'Average + Origin' approach for balancing efficiency with maintaining critical insights into sensory integration.

The performance of participants, measured by absolute error, did not differ significantly between the two tasks (Fig. 5). This result is further confirmed by the NASA-TLX results, which showed no significant differences in workload across the tasks. This consistency across tasks highlights the robustness of this cross-modal matching experiment and its potential applicability beyond the visual and haptic modalities explored here. Such an approach could be useful for studying other sensorimotor modalities where perceptual equity is critical, such as auditory-haptic or proprioceptive-visual interactions. Future work should build on this foundation by conducting similar experiments to explore these modalities, further broadening the scope and impact of cross-modal matching techniques in sensorimotor research.

Participant performance, measured by absolute error, was unaffected by the number of blocks in the experiment (Fig. 7). Performance remained consistent across blocks, indicating no significant changes due to fatigue or learning effects as the experiment progressed. This stability suggests that reducing the number of blocks, as proposed by the reduced order fit, would not negatively impact results. By maintaining consistent performance throughout the task, the reduced order approach ensures overall trends are captured accurately without compromising data integrity or missing potential learning effects. This finding further supports the feasibility of streamlining cross-modal matching experiments for improved efficiency.

In contrast, there was a clear effect of reference intensity on participants' error; participants exhibited greater errors at higher intensities, possibly due to sensory saturation (Fig. 6). This trend explains why the optimal reduced fits in the 'Average + Origin' dataset tended to exclude data points at higher intensities. This observation is a critical consideration for future experimenters. It suggests that certain ranges within a modality, in this case, higher intensities, may lead to diminished performance and weaker correlations between modalities. To address this, experiments should avoid operating within these problematic ranges and exclude them from cross-modal matching tasks to ensure robust results. Our formulation of reduced order fits offers a distinct advantage here: it programmatically identified these problematic intensity ranges and excluded them. Moreover, in stroke survivors, altered sensory processing may exacerbate sensory saturation, impacting clinical applicability. Future work should explore adaptive scaling corrections to optimize feedback precision.

A. Limitations

In this study, we initially compared linear and exponential fits to the data and found no statistical difference between the two fits. Based on this result and the predetermined understanding that the relationship between visual and haptic modalities in this task is inherently linear, we focused our analysis on linear fits. This approach was justified for this specific task, as the experimental design deliberately established a one-to-one relationship between modalities after normalization. However, other experiments may not have such a predetermined understanding of the underlying relationship between modalities. In these cases, it may be necessary to explore a wider range of models, including nonlinear fits, to accurately capture the relationship in the data. However, that is beyond the scope of this study.

Another limitation of this study is that it was conducted exclusively with healthy participants, which may limit generalizability to clinical populations, such as stroke survivors. However, our results showed that performance remained stable across blocks in healthy participants, with no evidence of fatigue affecting performance. This finding justifies the use of a reduced protocol, as reducing the number of blocks does not compromise data quality. Such a streamlined protocol is particularly valuable for patient populations, where fatigue and task duration are critical concerns. Additionally, stroke survivors' perceptual deficits may introduce non-linearity in matching tasks; here, we included exponential fits to account for such variations. Nonetheless, future research should validate these reduced protocols with clinical populations to ensure they retain effectiveness while accommodating the unique challenges posed by sensory and motor impairments.

B. Impact

This work advances cross-modal matching experiments by identifying reduced experimental designs that maintain accuracy while minimizing participant fatigue and time requirements. Here, we reduced the duration of the study by a factor of 5. These streamlined protocols are particularly beneficial in neurorehabilitation, where patients often face physical and cognitive limitations. By reducing the number of trials and reference intensities, this approach simplifies the experimental process while preserving data quality, paving the way for broader adoption in clinical and research settings.

The study also emphasizes the importance of perceptual equity in multimodal experiments by providing a framework for fair comparisons across sensory modalities. By aligning stimuli to achieve comparable intensities, it addresses a key challenge in sensory integration research [5]. These principles extend beyond visual and haptic modalities, offering a methodological foundation for investigating other sensory interactions and their roles in motor control and neurorehabilitation. For example, strategies such as limiting the upper stimulus range or implementing real-time adaptive algorithms could help mitigate errors in individuals with sensory deficits [18] and enhance clinical utility.

The methodological framework introduced in this study extends beyond just visual and haptic matching. This approach could complement protocols like those of Xu et al. [7] and Seim et al. [19], enabling perceptual equity calibration during individuated finger training. Integrating cross-modal matching into functional grasping or co-activation tasks could better tune sensory feedback across modalities, optimizing patient outcomes. By programmatically optimizing experimental parameters, this method scales to other sensorimotor modalities, such as auditory-haptic or proprioceptive-visual interactions. This scalability broadens its relevance to fields like prosthetics, virtual reality, and human-computer interaction, where multimodal integration is critical. Efficiently determining perceptual equity across modalities provides researchers with a versatile tool for advancing both experimental research and practical applications. Establishing perceptual equity through cross-modal matching is not just a refinement – it is a necessary evolution for multimodal research, rehabilitation, and beyond.

REFERENCES

[1] V. L. Feigin, M. Brainin, B. Norrving, S. Martins, R. L. Sacco, W. Hacke, M. Fisher, J. Pandian, and P. Lindsay, "World stroke organization (wso): Global stroke fact sheet 2022," *International Journal of Stroke*, vol. 17, no. 1, p. 18–29, Jan. 2022.

[2] R. B. King, "Quality of life after stroke," *Stroke*, vol. 27, no. 9, p. 1467–1472, Sep. 1996.

[3] B. Lee, I. D. Saragih, and S. O. Batubara, "Robotic arm use for upper limb rehabilitation after stroke: A systematic review and meta-analysis," *The Kaohsiung Journal of Medical Sciences*, vol. 39, no. 5, p. 435–445, Mar. 2023.

[4] J. M. Veerbeek, A. C. Langbroek-Amersfoort, E. E. H. van Wegen, C. G. M. Meskers, and G. Kwakkel, "Effects of robot-assisted therapy for the upper limb after stroke: A systematic review and meta-analysis," *Neurorehabilitation and Neural Repair*, vol. 31, no. 2, p. 107–121, Sep. 2016.

[5] C. M. Stinear, C. E. Lang, S. Zeiler, and W. D. Byblow, "Advances and challenges in stroke rehabilitation," *The Lancet Neurology*, vol. 19, no. 4, p. 348–360, Apr. 2020.

[6] J. Carducci, K. Olds, J. W. Krakauer, J. Xu, and J. D. Brown, "Novel planar strain sensor design for capturing 3-dimensional fingertip forces from patients affected by hand paralysis," *Sensors*, vol. 22, no. 19, p. 7441, Sep. 2022.

[7] J. Xu, T. Ma, S. Kumar, K. Olds, J. Brown, J. Carducci, A. Forrence, and J. Krakauer, "Loss of finger control complexity and intrusion of flexor biases are dissociable in finger individuation impairment after stroke," *eLife*, Nov. 2023.

[8] G. V. Pennati, J. Plantin, L. Carment, P. Roca, J.-C. Baron, E. Pavlova, J. Borg, and P. G. Lindberg, "Recovery and prediction of dynamic precision grip force control after stroke," *Stroke*, vol. 51, no. 3, p. 944–951, Mar. 2020.

[9] F. Abdollahi, E. D. Case Lazarro, M. Listenberger, R. V. Kenyon, M. Kovic, R. A. Bogey, D. Hedeker, B. D. Jovanovic, and J. L. Patton, "Error augmentation enhancing arm recovery in individuals with chronic stroke: A randomized crossover design," *Neurorehabilitation and Neural Repair*, vol. 28, no. 2, p. 120–128, Aug. 2013.

[10] J. A. Semrau, T. M. Herter, S. H. Scott, and S. P. Dukelow, "Vision of the upper limb fails to compensate for kinesthetic impairments in subacute stroke," *Cortex*, vol. 109, p. 245–259, Dec. 2018.

[11] A. Ahmed, B. Hugo, S. Lucas, R. Diana, O. Etienne, and G. Pascal, "Distinct and additive effects of visual and vibratory feedback for motor rehabilitation: an eeg study in healthy subjects," *Journal of NeuroEngineering and Rehabilitation*, vol. 21, no. 1, Sep. 2024.

[12] B. J. Pitts, S. A. Lu, and N. B. Sarter, "Cross-modal matching: The development and evaluation of a new technique," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 57, no. 1, p. 1760–1764, Sep. 2013.

[13] B. J. Pitts and N. B. Sarter, "Crossmodal matching: Validation of a more reliable technique," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 58, no. 1, p. 1751–1755, Sep. 2014.

[14] M. Paciaroni and M. Acciarresi, "Poststroke fatigue," *Stroke*, vol. 50, no. 7, p. 1927–1933, Jul. 2019.

[15] R. Oldfield, "The assessment and analysis of handedness: The edinburgh inventory," *Neuropsychologia*, vol. 9, no. 1, p. 97–113, Mar. 1971.

[16] S. G. Hart and L. E. Staveland, *Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research*. Elsevier, 1988, p. 139–183.

[17] N. Thomas, G. Ung, C. McGarvey, and J. D. Brown, "Comparison of vibrotactile and joint-torque feedback in a myoelectric upper-limb prosthesis," *Journal of NeuroEngineering and Rehabilitation*, vol. 16, no. 1, p. 70, jun 2019.

[18] C. Hazelton, K. McGill, P. Campbell, A. Todhunter-Brown, K. Thomson, D. J. Nicolson, J. D. Cheyne, C. Chung, L. Dorris, D. C. Gillespie, S. M. Hunter, and M. C. Brady, "Perceptual disorders after stroke: A scoping review of interventions," *Stroke*, vol. 53, no. 5, p. 1772–1787, May 2022.

[19] C. Seim, B. Chen, C. Han, D. Vacek, L. S. Wu, M. Lansberg, and A. Okamura, "Relief of post-stroke spasticity with acute vibrotactile stimulation: controlled crossover study of muscle and skin stimulus methods," *Frontiers in Human Neuroscience*, vol. 17, Aug. 2023.