Passive Haptic Rehearsal for Augmented Piano Learning in the Wild

TAN GEMICIOGLU, Georgia Institute of Technology, USA and Cornell University, USA
ELIJAH HOPPER, BRAHMI DWIVEDI, RICHA KULKARNI, ASHA BHANDARKAR, PRIYANKA RAJAN, NATHAN ENG, ADITHYA RAMANUJAM, CHARLES RAMEY, SCOTT M. GILLILAND, and CELESTE MASON, Georgia Institute of Technology, USA
CAITLYN SEIM, Stanford University, USA
THAD STARNER, Georgia Institute of Technology, USA

Fig. 1. a) Haptic gloves for daily passive haptic rehearsal. b) Practicing piano with our piano education system and a Casio keyboard. c) Online learning portal of our piano education system.

Passive haptic learning (PHL) is a method for training motor skills via intensive repetition of haptic stimuli while a user is focused on other tasks. For the practical application of PHL to music education, we propose passive haptic rehearsal (PHR) where PHL is combined with deliberate active practice. We designed a piano teaching system that includes haptic gloves compatible with daily wear, a Casio keyboard with light-up keys, and an online learning portal that enables users to track performance, choose lessons, and connect with the gloves and keyboard. We conducted a longitudinal two-week study in the wild, where 36 participants with musical experience learned to play two piano songs with and without PHR. For 20 participants with complete and valid data, we found that PHR boosted the learning rate for the matching accuracy by 49.7% but did not have a significant effect on learning the notes' rhythm. Participants across all skill levels in the study require approximately two days less to reach mastery on the songs practiced when using PHR. We also confirmed that PHR boosts recall between active practice sessions. We hope that our results and system will enable the deployment of PHL beyond the laboratory.

CCS Concepts: • **Human-centered computing → Haptic devices**; **Empirical studies in ubiquitous and mobile computing**; **Ubiquitous and mobile devices**.

Additional Key Words and Phrases: haptic, tactile, wearable, passive training, PHL, piano, learning, rehearsal

Authors' addresses: Tan Gemicioglu, tg399@cornell.edu, Georgia Institute of Technology, Atlanta, GA, USA and Cornell University, New York, NY, USA; Elijah Hopper, ehopper3@gatech.edu; Brahmi Dwivedi, bdwivedi6@gatech.edu; Richa Kulkarni, rkulkarni83@gatech.edu; Asha Bhandarkar, abhandarkar6@gatech.edu; Priyanka Rajan, prajan31@gatech.edu; Nathan Eng, neng32@gatech.edu; Adithya Ramanujam, adithyar697@gmail.com; Charles Ramey, cramey7@gatech.edu; Scott M. Gilliland, scott.gilliland@gatech.edu; Celeste Mason, celeste.m@gatech.edu, Georgia Institute of Technology, Atlanta, GA, USA; Caitlyn Seim, cseim@stanford.edu, Stanford University, Stanford, CA, USA; Thad Starner, thad@gatech.edu, Georgia Institute of Technology, Atlanta, GA, USA.

2474-9567/2024/12-ART187
https://doi.org/10.1145/3699748

ACM Reference Format:
Tan Gemicioglu, Elijah Hopper, Brahmi Dwivedi, Richa Kulkarni, Asha Bhandarkar, Priyanka Rajan, Nathan Eng, Adithya Ramanujam, Charles Ramey, Scott M. Gilliland, Celeste Mason, Caitlyn Seim, and Thad Starner. 2024. Passive Haptic Rehearsal for Augmented Piano Learning in the Wild. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 4, Article 187 (December 2024), 26 pages. https://doi.org/10.1145/3699748

1 INTRODUCTION

Learning a new instrument is a complex task that requires learning musical notation, practicing fine motor coordination, and learning particular musical pieces to perform. Sustaining musical practice is even harder: 50% of musical students quit learning by the time they are 17 years old [41] and only 9.6% of US adults play a musical instrument regularly [31].

Passive haptic learning (PHL) is a technique for learning motor skills via repeated haptic training even when little to no attention is dedicated to learning [47]. PHL uses ambient tactile cues as instructions to entrain sequential movements and associated information. PHL has been an effective approach for training a multitude of skills starting with piano [18], and including Braille [50], Morse code [48] and stenography [1]. More recently, PHL for piano has been shown to provide benefits to retention [8] and work with modalities beyond vibrotactile haptics [12]. However, past studies of PHL have been conducted in the lab, and studies of its longitudinal effects have been limited in the piano domain. These studies prioritized internal validity, focusing on narrow conditions: simple, short songs, rigid pedagogical methods, and inexperienced learners.

Learning has little purpose when the practiced skill is never used. As such, it is important to consider how active musical practice and PHL influence each other. For PHL to be successful as a tool in musical practice, it must be usable in the long term by real musical students practicing complicated songs and engaging in multiple forms of practice based on personal preferences. We propose a hybrid piano learning technique integrating passive training with deliberate active practice to reinforce developing skills, boost recall, and accelerate learning. In contrast with past research on PHL which has focused on piano practice purely using haptic stimulation, we call our new hybrid learning approach passive haptic rehearsal (PHR).

To evaluate the feasibility and effectiveness of passive haptic rehearsal, we conducted the first longitudinal in-the-wild study of passive haptic learning, where 36 participants with musical experience practiced two songs – with and without passive rehearsal – over the course of two weeks, in their own homes. 20 participants fully completed the study with valid data. In this study, we aimed to answer two main research questions:

(1) How much does daily PHR improve the learning rate of practicing a new song in-the-wild?
(2) Does PHR help students remember how to play a piece between active learning sessions?

To support the study, we designed haptic gloves that were suitable for daily wear and an online learning portal that could connect with the haptic glove for starting rehearsal sessions. The portal could also connect with a Casio keyboard with light-up keys for recording MIDI files of a student's performances and give the participant feedback on their learning progress. Participants recorded their performance before and after their active practice session every day, which we used to study the learning curves of their progress over time.

We make the following contributions:

- To the best of our knowledge, the first in-the-wild study of passive haptic learning for piano and the characterization of its effect on the learning curve.
- A new hybrid approach that we call passive haptic rehearsal, combining passive haptic learning with spaced active practice, and pedagogical insights into its application.
- An open-source system integrating recent advances in piano practice compatible with MusicXML and MIDI keyboards to enable passive rehearsal for any musical piece.[1]

[1]https://github.com/tangemicioglu/phl-portal

2 RELATED WORK

The design of our study, learning system, and haptic glove are built on the past decade of research on passive haptic learning. We also expand further to gain insights from the longer history of musical learning and recent interfaces designed to advance musical learning. We combine them to optimize the learning benefits that self-paced piano learners may gain from passive haptic rehearsal.

2.1 Passive Haptic Learning

Passive learning is "caught, rather than taught" [22] and facilitates learning even in the absence of motivation and effort [57]. The application of passive learning to haptics started with Huang et al.'s work on Mobile Music Touch [17, 18]. Huang et al.'s studies used a variety of high cognitive load tasks to evaluate whether people learned piano when they received tactile stimuli on their fingers even when their attention was occupied by other tasks and observed that people were able to learn to play simple sequences without any active practice. The phenomenon came to be known as passive haptic learning (PHL) [18, 21]. PHL uses tactile stimulation to teach motor skills passively via repetitive instructional cues applied to the fingers [47]. Markow evaluated PHL without the use of auditory stimuli, showing that haptics alone was sufficient for learning [27]. While early studies tested simple melodic "phrases" on one hand, later work gradually expanded the usage of PHL to two-handed, chorded piano pieces and confirmed that it was still effective [47]. These studies have proven PHL to be robust, and shown it persists even during more complex applications.

Further studies found additional effects of PHL when learning piano. Donchev et al. found that PHL was as effective as active practice in unaided recall, and more effective than active practice when supported by audiovisual cues before performing [8]. Donchev et al.'s work shows PHL can be very useful for piano education, as many learners have limited time for regular active practice. Fang et al. expanded the application of PHL to non-vibrotactile haptic modalities, showing that stroking or tapping sensations could enable equivalent or greater learning. Another study by Fang et al. evaluated how PHL affected participants' ability to perform rhythmically based on relative timing and duration of musical notes [13]. Their study found that PHL could enable improved rhythm, although combining audio and haptics provided greater outcomes. We also decided to test the impact of PHL on rhythm as it is an important component for performing musical pieces correctly.

Many studies have also developed alternative applications of PHL and aimed to develop an understanding of other effects it may have. Seim et al. conducted several studies successively, demonstrating the effects of PHL in teaching Braille [43, 50], Morse code [49] and typing [46]. Seim et al. also demonstrated PHL on several commercial wearables such as Google Glass [49] or a Sony Smartwatch [48], giving valuable design ideas towards more robust PHL hardware. Luzhnica et al. expanded PHL to skin reading [25, 26] while Aveni et al. expanded it to stenography [1]. Recent work has shown that the principle behind PHL can also be used for rehabilitation [11, 45, 51] and performs better than the existing clinical practice for rehabilitating post-stroke spasticity [44]. Combined, these works form a portrait of PHL as working across a diverse range of modalities and devices, enabling it to be suitable for everyday use.

Notably, all of the past studies of PHL were conducted within the lab. Aside from Donchev et al.'s work on retention in piano [8], they used a single session or a series of sessions within a short duration. Thus, there is a significant gap in how passive haptic learning may perform in the wild and over a longer duration.

2.2 Musical Learning and Interfaces

Learning and practicing music can provide a wide range of cognitive benefits across different age groups, such as improving intelligence and working memory in adolescents [3] and benefiting mental health, cognitive abilities, and motor function in older adults [29]. However, sustaining musical learning requires significant effort. The greatest predictor of musical success is practicing regularly and consistently [52]. Learners benefit most when they engage in deliberate practice [10] and use spaced repetition to enhance the retention of their new skills [20]. On the other hand, consistent feedback can encourage learners' progress and prevent dropping out of learning [5]. Existing work in piano skill acquisition primarily focuses on learning assisted by a piano instructor or school [5, 52]. Adults learning music often have a unique context, where they must have self-direction and autonomy [40]. We particularly focused on adult learners who were engaging in self-paced practice, as they have the most to gain from additional feedback systems and passive haptic learning.

The human-computer interaction community has designed a multitude of tools and musical interfaces for assisting self-paced learners. Some musical interfaces provide more incentives as the learner spends time practicing piano, ensuring continuity and regular practice [37]. New piano interfaces use visual cues, such as light-up keys, new piano notations [39], and finger position tracking [28] to improve performance and reduce the cognitive load demanded by practice. Meanwhile, other systems attempt to teach people to perform rhythmically [53] or allow more expressiveness [19]. Several prototypes have also attempted to aid musical learning by enabling synchronization of movements with a teacher [34, 56] or by directly actuating movements towards the manner they should be performed [33, 54]. A systematic review of augmented piano systems by Deja et al. [7] found that while the synchronization space is saturated, more work is needed for encouraging sight-reading, motivation, and improvisation. While assisting with improvisation or sight-reading was out of scope for this project, we attempted to improve motivation using passive haptic rehearsal and continuous feedback. We also let participants use audiovisual guidance via light-up keys and auto-play during their practice sessions.

For the purpose of our study, it's important to be able to model musical learning over time in addition to assisting it. Research on cognition and music has used a wide range of models for performance over time. Musical skill acquisition has been modeled based on a power law function in the past [9], but later work on the law of practice suggests individualized models fit an exponential function better [16]. Recent work has continued to use the exponential function for modeling the law of practice and found it to be an appropriate approach for depicting the learning of motor skills [35, 38]. While there have also been recent work suggesting the use of S-curves or sigmoid functions [24], we chose not to test such a model because its complexity compared to our current number of participants made a fit challenging. For the sake of validation, we compared a linear, power and exponential function as further described in Section 5.1.

3 DESIGN

3.1 Passive Haptic Learning Gloves

The PHL gloves are each composed of three subsystems: a cable assembly with connected vibrotactile motors, a microcontroller and battery housed in a case, and an off-the-shelf faux leather glove. The five vibrotactile motors are positioned at the base of each finger on the proximal phalanxes and interior surface of the glove. To avoid direct contact with the motor which may cause discomfort or allergic irritation due to the metals, the motors are covered by a thin layer of polyester between the skin. The motors are connected to the microcontroller and battery housed within a 3D-printed case (shown in Figure 2) located at the back of the hand, protecting the electronics against sweat, dust, and mechanical wear and tear.

The gloves use Precision Microdrives' 310-103 coin-shaped eccentric rotating mass (ERM) vibration motors, consistent with past PHL studies using similar motors by Precision Microdrives [8, 47]. The motors imitate the order in which the piano keys are played and repeatedly stimulate the user's fingers used to play those keys. While active, the motors are driven at 3.3V and vibrate with a constant amplitude of 1.5G and frequency of 240 Hz until the end of the note. Further details on the ordering and tempo are described in Section 3.2.

Users can connect to the gloves over Bluetooth Low Energy (BLE) through a web application and select which piano piece to deploy to the gloves. The gloves save the selected song on-board, allowing them to operate independently and eliminating the need for a continuous external device connection which could otherwise

Fig. 2. Apparatus taken home by participants during study, including custom haptic gloves, Samsung Galaxy Tab tablet and Casiotone LK-S250 piano keyboard.

limit daily activities. The right-hand glove controls the left via BLE for synchronization of both hands during the passive learning session. Furthermore, four status LEDs are mounted on the case to allow progress tracking during the session.

Inside the case, a rigid PCB utilizing the TinyPico Nano development board and a 1S 290mAh LiPo battery are mounted. All the required user interface components, including the USB connector for charging, LEDs, and power and pause buttons, are directly mounted onto the PCB with openings through the case for user access. The 290mAh battery capacity provides approximately three hours of passive haptic learning– sufficient for one passive learning session. The protective case and the motors are fitted on a pre-made synthetic leather glove for a snug fit and flexibility. The fingerless design of the gloves allows users to perform their daily tasks without interference.

3.2 Stimulation Pattern

During passive haptic learning sessions, the passive haptic learning gloves activate the vibration motors against participants' fingers according to the fingerings of the song they are learning. Every note of each song has a corresponding suggested finger used to play that note on the piano. The PHL gloves play through the song, delivering haptic stimulation to the participant's corresponding fingers as the song plays. Each note's corresponding motor is vibrated at constant amplitude for the entirety of its musical length with a *allegro* tempo of 120 beats per minute (e.g., a quarter note would be vibrated for 500ms), equivalent to the songs' tempo during practice. For chords, the motor for each note is activated concurrently. Haptic stimulation between notes is spaced with a 100ms pause to provide a clear sense of separation between notes. The haptic stimulation is further chunked into phrases, consistent with past PHL stimulation paradigms [4, 42]. Each phrase is generated by incrementally including measures (with a 2/4 time signature, one measure is equal to a half note or four eighth notes) until at least ten notes are in the phrase. A 2-second pause clearly signals the end of a phrase. Phrases are repeated 10 times each before the gloves proceed to the next phrase. After the final phrase in the song is played, the stimulation returns to the first phrase to begin a new cycle. The entire song is repeated in this manner until two hours have elapsed.

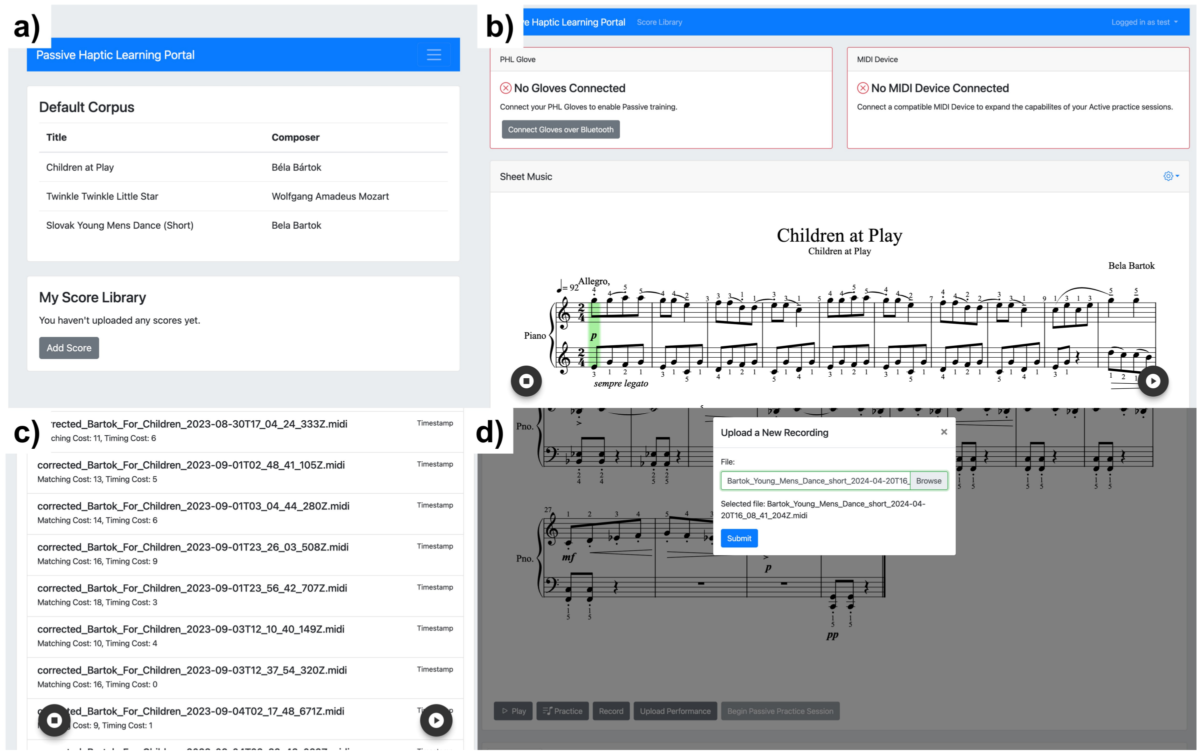

Fig. 3. Learning portal, showing: a) Homepage for selecting song or uploading new songs. b) Song page, where gloves or MIDI keyboard can be connected to portal. c) Cost section of song page, showing past recorded performances. d) MIDI recording upload for cost measurement.

3.3 Online Learning Portal

We developed an online learning portal using a progressive web app to track learners' progress over time and to enhance the usability of the system (see Figure 3). The user accesses the online learning portal with a user ID and password and then chooses a song to learn from a selection list. Typically, users have access to a study curriculum and other songs from a public corpus. The user views the sheet music for the selected song augmented with the fingerings labeled for each note for both hands.

The fingering data is necessary to produce the vibrotactile stimulation patterns for the PHL gloves. However, such fingerings are typically not available from commonly available MusicXML scores [15]. Thus, when a new MusicXML score is uploaded to the site, we derive the fingerings semi-automatically using the *pianoplayer* package [14, 30]. While manual tweaks are still necessary to tune the automatically produced fingerings, we seek to fully automate the process in the future.

The portal also allows the user to view the Bluetooth connectivity status of the gloves and the USB connection to the MIDI keyboard. When the user presses the 'Play' button, the song audio plays with a cursor moving along the sheet music to the corresponding notes. The users have an option for the corresponding keys on the MIDI keyboard to light as the song plays, and the tempo at which the song plays can be increased or decreased.

The user can press the record button to record themselves performing the piece on the MIDI keyboard. The user can then upload these MIDI files to the web portal where they can view an evaluation of their performance.

After multiple performances, users can view and track their improvement over time. The portal also allows users to enter personal notes on their practice. Our software makes heavy usage of the computational musicology library *music21* for processing and interacting with digital scores [6].

4 METHODS

4.1 Participants

Our study was approved by the institutional IRB board at the Georgia Institute of Technology. We recruited 36 participants to complete a two-week in-the-wild study on augmenting regular piano practice with passive haptic rehearsal. All participants had a musical background either in piano or another instrument. We primarily recruited participants through word of mouth and email lists. We required that participant be 18 to 75 years old, have enough sensory perception to feel vibrations on their hands, have no pre-existing medical conditions that would prevent piano practice, not be undergoing therapy or rehabilitation for upper extremity disorders, be able to read sheet music, and possess piano skills that were not too advanced. We considered "too advanced" to be when participants were able to play the songs used in the study with less than ten mistakes during their baseline recording. "Too advanced" participants perfected each song within one day, resulting in no significant learning data for the rest of the week. Participants were classified as possessing either beginner, intermediate, or advanced piano skills as described in Section 4.4. Due to the long total duration of the study (3 lab visits, 7 hours of active practice, and 14 hours of passive haptic rehearsal), participants each received $140 for their participation.

During the study, eight participants had to be excluded due to violating one of the requirements: Three participants had a gap in practice that was longer than a week, two demonstrated no learning during their first week, two were too advanced for the study, and one was discovered to have not known sheet music during the study. During analysis, we additionally discarded eight more participants, as the first five participants had inconsistent protocol and three participants were affected by a MIDI recording glitch making the majority of their data invalid. For participants with less than half of their data missing (due to the recording glitch or missed recording sessions), we included them with empty data points for missing records, as our analysis tools were capable of handling missing data as described in Section 4.5. Therefore, we only used data from 20 participants during analysis. Among those 20 participants, 6 identified as female, 13 as male, and 1 as non-binary. Regarding ethnicity, 13 identified as Asian, 1 as Black, and 6 as White. Participants were between 18 and 26 years old, with an average age of 21.15 years and a standard deviation of 1.85 years. 8 participants were beginners, 5 were intermediate, and 7 were advanced. All participants had experience with either piano or another instrument, ranging from one year to twelve years. Participants had 1.95 years of experience on average for piano and 4.6 years for other instruments. The detailed musical background of each participant is shown in Appendix A. The 20 participants were recruited between February 2023 and February 2024.

4.2 Study Design

We ran a mixed-subjects study with conditions balanced according to a 2x2 Latin square to mitigate ordering effects. Each participant completed the study over the course of two weeks, practicing two songs by the Hungarian composer Béla Bartók: Children at Play (CAP) and Slovak Young Men's Dance (SYMD). The two songs were close to equal in number of notes and had similar combinations of eight and quarter notes, amount of chords, usage of each hand, and were generally at the same level of difficulty. Complete sheet music with finger numbering for these songs is shown in Appendix B. For the participants' first week, they were assigned to either the control or stimulus condition, and one of the two songs to practice. Five participants started week 1 with the stimulus condition and Children At Play song, seven with control and Children At Play, three with Slovak Young Men's Dance, and five with control and Slovak Young Men's Dance. During the second week, each participant was assigned the condition and song they were not assigned for the first week. As participants were recruited, they

Original Times:	t_1	t_2	-	t_3	t_4	-	t_5	t_6	-
Original Notes:	C	B	B	A	E	E	F	C	-
Performed Notes:	C	B	-	A	E	C	F	C	C
Performed Times:	t'_1	t'_2	-	t'_3	t'_4	-	t'_5	t'_6	-
			Deletion			Substitution			Insertion

Fig. 4. The notes played during a recording being compared to the original song notes using the Needleman-Wunsch algorithm.

were internally paired with another participant of the same skill level and song order but in opposite experimental conditions to maintain counter-balancing.

Each day of the study, participants used their tablet and keyboard at home to record an attempt of performing their assigned song. After recording, they were instructed to practice that song on their keyboard for a total of 30 minutes. Afterwards, they used the tablet and keyboard to record another attempt at playing that song. They uploaded all their recordings to our online learning portal and noted the outputted match cost and time cost. During the stimulus condition week, participants also wore the passive haptic learning gloves for two hours each day before recording their first attempt at the song. The gloves were turned on and actively delivered haptic stimulation throughout the entire two hours. In this way, the passive haptic learning gloves augmented traditional piano practice for the stimulus condition week.

At the start and end of each week, participants came to the lab to perform and submit baseline recordings, receive study materials and instructions, and participate in interviews. Initial lab visits involved sight-reading and recording the assigned song for that week, while final visits required participants to record their last attempt at the song and answer additional questions about their study experience and regular practice routines. All equipment was returned at the end of the study, and participants were compensated for their involvement.

To avoid some risk of negating the effects of passive haptic learning, we instructed participants to not listen to music, sleep, or participate in any intense activities while wearing the gloves. We did allow participants to break practice into separate chunks. For example, wearing the gloves for an hour, taking a two hour break, then wearing the gloves for another hour was permissible. Finally, we prohibited participants from receiving any external help with the songs, from a piano teacher or otherwise.

4.3 Measures

The online learning portal has an evaluation feature which automatically processes users' performances as MIDI files and outputs two metrics: a "match cost" and a "time cost". The match cost is a measure of how many correct notes the user did not play plus how many wrong notes the user did play. The time cost is a measure of how many notes the user played outside of the correct time by a threshold T. Key presses that occurred at approximately the same time are treated as chords, and are processed as unordered sets during the calculation of the match cost.

The match cost was calculated with the Needleman–Wunsch algorithm [32] as described by Gemicioglu et al. [14]. This dynamic programming algorithm is "stricter" when assigning error costs as compared to the previously, and much more commonly, used sequence alignment algorithm, Dynamic Time Warping, which ignores erroneous recurring presses of the same key [18, 47]. An example of such repeated notes, which occur in many songs, and how the Needleman-Wunsch algorithm identifies them is shown in Figure 4. The measurement of rhythmic inaccuracies was incorporated into the algorithm using a timing error based on whether the difference between user-played and original (based on the sheet music) key presses in matched sequences were above a threshold T. If the user-played timing was too far from the original, then it was considered an error, and time cost was incremented. In our analysis, we also considered the "total cost", the sum of both costs.

$$\text{Time Cost} = \sum_{i=0}^{m} 1(|t_i - t_i'| \geq T)$$

$$\text{Match Cost} = \text{Deletion cost} + \text{Insertion cost} + \text{Substitution cost}$$

$$\text{Total Cost} = \text{Match Cost} + \text{Time Cost}$$

The match, time and total cost were measured sixteen times during each week of the study: Once for a initial baseline recording, seven times for daily pre-practice recordings, seven times for daily post-practice recordings and once for a final evaluation. For statistical evaluation, we normalized the cost metrics by the number of notes in each song, balancing out a minor difference in length between CAP (148 notes) and SYMD (165 notes). Experimenters observed that when participants made less than or equal to three mistakes, their performance was indistinguishable from being perfect. We converted this threshold to the normalized form, considering participants to have mastered the song if their matching cost was less than 2%. The estimated duration to reach this level was considered the "days needed to reach mastery."

Using the normalized cost, we also built additional metrics for statistical evaluation. The difference between each successive post-practice session showed the effect of daily practice, comparable with post-session evaluations in past PHL studies. The difference between baseline and evaluation showed the longitudinal effect over the duration of the entire study. Meanwhile, for evaluating per-session active practice and passive haptic rehearsal independently, we relied on the difference between pre-practice and post-practice evaluations. We considered the difference between post-practice and pre-practice on the same day to be due to active practice, while the difference between post-practice from the previous day and pre-practice on the next was due to retention or passive haptic rehearsal.

4.4 Procedure

After providing written informed consent, participants completed a demographic questionnaire and were randomly assigned to one of the four possible experiment condition orders (as explained in Section 4.2). Participants then listened to their assigned song for the first time and recorded a sight-reading attempt to the best of their ability to serve as a baseline with no practice. The study session was conducted in a quiet research study area with study personnel only.

As shown in Figure 2, participants were given a Casio LK-S250 61-Key Portable Keyboard, a Samsung Galaxy Tab A8 tablet, and a pair of passive haptic learning gloves to take home for the duration of the two-week study. We also provided an instructional packet with information on all the provided materials. Participants were instructed to practice for thirty minutes in the manner they prefer (e.g. using audiovisual playback, following guidance from light-up keys, or simply practicing by following the sheet music) and submit recordings as specified in their packets. Participants' daily schedules is outlined in Section 4.2 and detailed in Figure 5. Participants were instructed to upload each recording to the online learning portal on the tablet and write the match and time cost outputted by the online learning portal in the corresponding blank in their instructional packet.

One week into their study, participants were required to come into the lab to switch songs and experimental conditions as well as answer interview questions. To switch songs and conditions, participants recorded a final attempt at their first week song, listened to their next song for the first time, and recorded a first sign reading of the new song for their baseline. When participants had finished their control condition week, we asked them if they faced any problems with the UI of the online learning portal and if they had any suggestions to improve it,

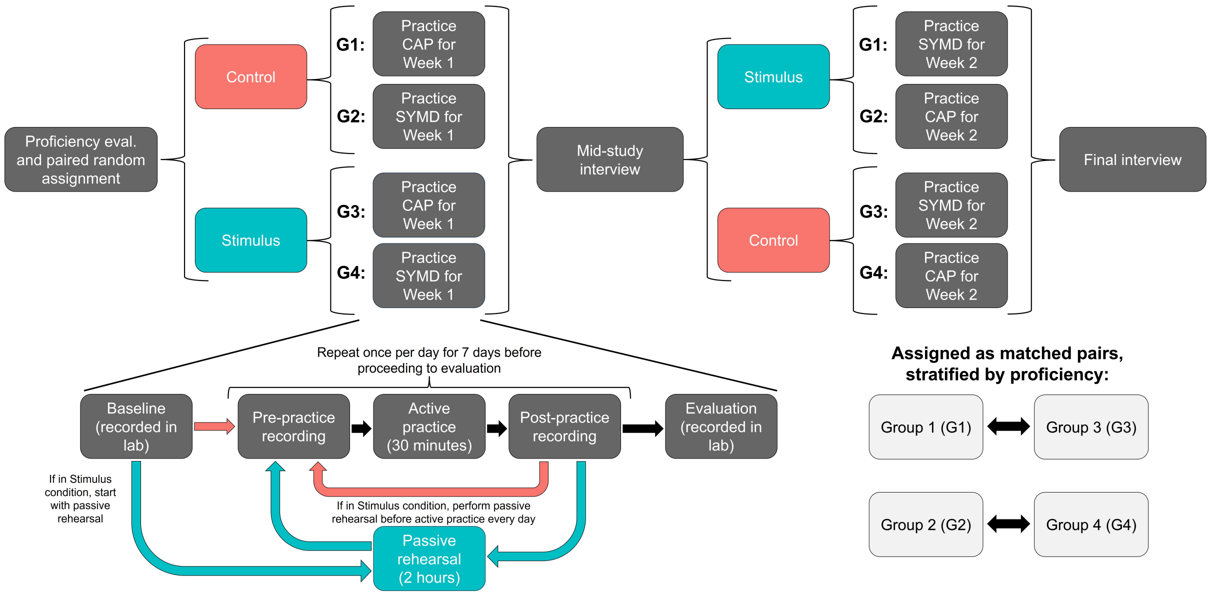

Fig. 5. Flowchart of study procedure and counter-balanced groups, including the weekly recording and practice schedule of participants. Each participant was assigned to one group from G1 to G4 with their matched pair (shown in bottom right corner), practicing the songs with PHR (stimulus) or without (control) for each week. Arrows marked in blue only apply to stimulus condition while arrows marked in red only apply to control condition.

if they found their assigned song easy or difficult to learn, and if they faced any issues with the study system as a whole. When participants had just finished their stimulus condition week, we asked them the same questions about the UI and how easy or difficult learning the song was, as well as how comfortable the gloves were to wear, if the gloves hindered day-to-day activities, and what they would change about the gloves.

Upon the conclusion of the two week study, participants came to the lab to submit a final recording, answer interview questions, and return all study material back to the lab. In addition to the corresponding questions mentioned earlier, participants were asked which song they found more difficult to learn, whether they felt that the gloves helped them learn the song faster, and how they normally go about learning and practicing a new song.

4.5 Analysis and Modeling

All analyses were conducted using *R Statistical Software (v4.3.2)* [36]. We used linear mixed-effects models (LMEs) to model data and conduct significance tests. Due to imbalances in our data, it was important to remove the effects of different factors from the effect of condition. LMEs were capable of handling the imbalance between different conditions and factor levels, as well as missing data due to participants who missed recording sessions. We investigated the complex interactions of participants' matching and timing costs with days of practice (1 to 7), condition (control without PHR, stimulus with PHR), proficiency (beginner, intermediate, advanced), block (1, 2), and music (Children at Play, Young Men's Dance). All LMEs were constructed using the *lme4* package of R [2], and statistical significance tests were conducted using the *lmerTest* package [23].

Our analysis of the learning curve proceeded in three stages: First, we attempted to determine the relationship of cost with days of practice by constructing non-linear mixed-effects models for the cost over time as linear

($\boldsymbol{Cost = a \cdot t + k}$), as a power law ($\boldsymbol{Cost = a \cdot t^{-b} + k}$), or as exponential ($\boldsymbol{Cost = a \cdot e^{-b \cdot t} + k}$). Next, we linearized the best model from the previous stage using the natural logarithm and tested different models by gradually increasing the number of terms from factors other than the experimental condition. We included the experimental condition and its interaction with day in each model as it would not be possible to determine its effects if it was not in the model. Finally, after selecting the best model, we used it to compare the overall learning between post-practice evaluations in matching, timing, and total cost.

To determine the best models in the first two stages of our analysis, we constructed each model, conducted ANOVA (analysis of variance) tests between the models, and compared their Akaike Information Criterion (AIC) – a common and effective metric for determining the relative quality of different statistical models [55]. We incrementally compared the models in pairs, using the best model (the one with lower AIC) to compare with the next model in the list. Where no significance was found, we preferred the simpler model or the one advocated by prior literature. All tests for these stages were conducted using the matching cost and post-practice tests, as these were the metrics used across past studies of PHL [8, 47]. We did not repeat the model selection process for the time cost and total cost metrics to ensure consistency in comparisons.

We also analyzed the effects of daily practice, baseline vs. evaluation, per-session active practice, and per-session passive haptic rehearsal using LMEs. More information about calculating each metric can be found in Section 4.3. The effect of daily practice was a simplified version of the learning curve model for comparability with past work, although we got the difference in post-practice costs between each day instead of each session [47, 49]. The baseline vs. evaluation metric was used to assess longitudinal improvement over the course of the entire study section. The per-session active practice and PHR metrics provided valuable insights into the independent effects of active practice and PHR without the direct influence of the other. As these were differences in cost instead of the raw cost values, they did not yield learning rate information. Thus, the interactions of factors with day were excluded in their design.

5 RESULTS

In this section, we constructed the learning curve of PHR using a multi-factor LME model. We found that PHR improves the learning rate by 49.7%, enabling participants to master a new song two days earlier. We also found that PHR not only eliminates the forgetting effect between active practice sessions, but allows learning to continue without practicing. Section 5.1 describes how the model is selected and the effects of different factors on learning and performance. Section 5.2.1 isolates the effect of passive rehearsal and active practice on recall and learning. Section 5.2.2 evaluates daily and weekly learning for comparability with past PHL studies.

5.1 Learning Curve

5.1.1 Model Selection. We first started to model the learning curve by determining the relationship between the cost or error of participants' performances and days of practice. Moving from a linear model to a power law significantly improved model fit ($\Delta AIC = 23.9$, $\chi^2 = 25.872, p < 0.001$). The exponential model was marginally better than the power law model ($\Delta AIC = 0.6$, $\chi^2 = 0.631$) but could not be tested for significance since they had the same number of terms, as described in Section 4.5. Neither model was simpler, so we chose the exponential model as it was well-supported by past literature in motor learning [16, 38] and had the lowest AIC.

While the exponential function for learning is naturally non-linear, it can be transformed to a linear form by removing the k term and taking the natural logarithm of both sides: $\boldsymbol{\log(Cost) = log(a) - bt}$. This equation is more suitable for linear approximation, but with an LME, it makes the assumption that the *initial performance* of cost (a) is linearly normally distributed. We instead assume that a is logarithmically normally distributed using the equation $\boldsymbol{\log(Cost) = a - bt}$ where a is *initial performance*, b is *learning rate* and t is *days of practice*, commonly known as a log-linear model. The log-linear expression of cost over time is a clear fit to the average

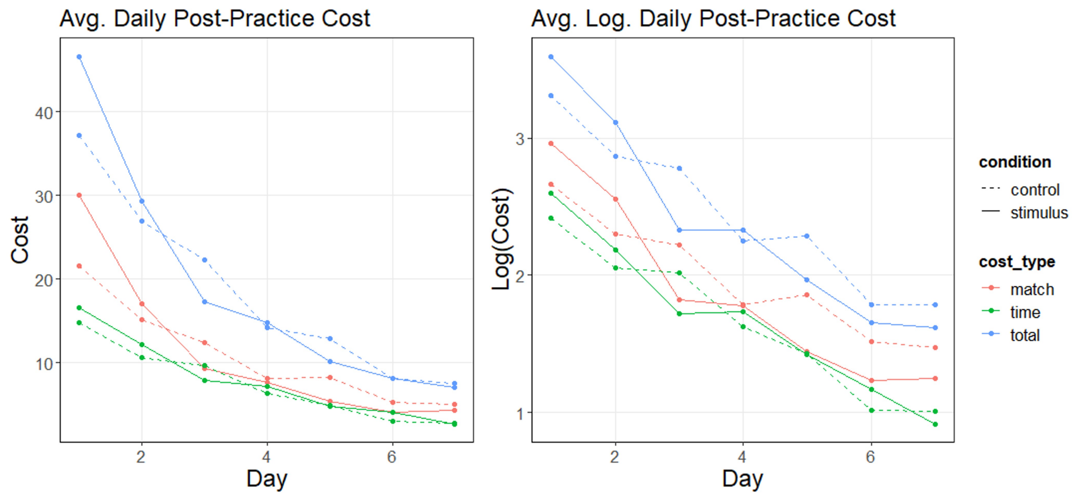

Fig. 6. Average cost (error) between all participants (n=20) for each day of practice, by condition and cost type. Lower cost indicates better performance. Left plot shows cost on a linear scale while the right plot shows it on a logarithmic scale. Cost over time on the right plot is close to a linear relationship, demonstrating the log-linear progression. However, many factors are averaged out, making the effect of PHR difficult to interpret and necessitating a multi-factor statistical model.

cost per day between participants and makes it more interpretable, as shown in Figure 6. However, the averaging of different factors in the plot makes the effect of PHR difficult to interpret, necessitating a multi-factor LME for analysis.

Using our linearized model, we then constructed three LMEs. The first model only used the effect of time for overall learning rate, condition (control or stimulus) on initial performance and learning rate, and a random effect on initial performance for the specific participant. The second model added the effects of musical piece (CAP or SYMD), block (1 or 2) and proficiency (beginner, intermediate, advanced) on initial performance. The third model added the effects of musical piece, block and proficiency on the learning rate. We did not build models that were more complicated than the third model – such as evaluating the interactions of condition and proficiency over time – because there would be too many terms (less than 20 data points per term) to approximate for the amount of data we had.

We first compared the first model with the second model and found that the second model was significantly better at fitting the data ($\Delta AIC = 8.61, \chi^2(4) = 16.606, p = 0.002$), indicating that the additional factors had an important role in the initial performance of participants. Then, we compared the second model to the third model and found that the third model was significant better than the prior models at modeling the data ($\Delta AIC = 1.36, \chi^2(4) = 9.361, p = 0.05$), indicating the additional factors had a role in learning rate, but a relatively minor one based on the low AIC difference. Therefore, we selected the third model, including the effect of each factor on the initial performance and learning rate. The importance of these factors on performance and learning was unsurprising, particularly because participants may learn at different rates based on their proficiency. Beginners in the study never achieved mastery within the one-week duration, whereas some advanced participants already had a perfect score by the third day. The individual participant data is shown in Appendix C, demonstrating the influence of proficiency and the variance between participants. Our final chosen model can be expressed as an equation of the form:

$$\log(Cost) = t \cdot \sum_{i=0}^{5}(\beta_i x_i) + \sum_{i=1}^{5}(\alpha_i x_i) + \epsilon$$

where t is days of practice, $x = (1, Condition, Music, Block, Proficiency)$ is the levels of each factor for a given participant, β contains the regression coefficients for the interaction of each factor with time (learning rate), α contains the regression coefficients for the main effect (initial performance) and ϵ denotes the random effect due to the participant.

5.1.2 Effect on Learning and Performance. We developed a model for each of logarithmic match cost, logarithmic time cost and logarithmic total cost across participants' daily post-practice recordings. For the factors, we took the control condition, CAP music, and beginner proficiency to be the reference level for comparisons. Block and day were treated as numeric and did not need a reference level.

For the match cost, we found that only the block (t=-1.98, p=0.049) and proficiency (Intermediate: t=-2.50, p=0.019; Advanced: t=-4.12, p<0.001) had a significant effect on participants' initial performance. Participants that were in their second week and participants with more piano experience made less mistakes. As for the learning rate with match cost, we found that both condition (t=-2.5, p=0.013) and advanced proficiency (Intermediate: t=1.3, p=0.19; Advanced: t=2.70, p=0.007) had a significant effect on participants' learning rate. Therefore, participants were *learning significantly faster* when they were using PHR, while participants who were experienced appeared to be learning slower as they had fewer mistakes to fix compared to other conditions. Overall, these results suggest that PHR is successful for improving the ability of participants to learn the correct sequence of notes, but proficiency is an important consideration for learning.

We then modeled and tested the time cost in the same manner. We found that the musical piece being practiced (t=2.88, p=0.004) and proficiency (Intermediate: t=-2.69, p=0.013; Advanced: t=-4.56, p<0.001) had a significant effect on the participants' initial performance. The pattern of proficiency's importance in ability was consistent with match cost, but the finding that musical piece affected the rhythmic accuracy had interesting implications. This effect may have emerged due to the different *pacing* of CAP and SYMD, as SYMD has a multitude of rests between notes. The learning rate had a significant effect only due to proficiency (Intermediate: t=3.38, p=0.019; Advanced: t=-4.12, p<0.001). Once again, participants with more experience made less mistakes overall, thus having less mistakes to fix. Since the impact of condition was not significant (t=0.099, p=0.92), we can not conclude that PHR affects the ability to perform rhythmically. We discuss why this effect may have differed from Fang et al.'s study [13] further in Section 6.4.

We also tested the total cost to see if there were combined effects, but we note that a composite metric based on both match and time cost may have a less-clear pattern relating to the performance. The music piece being practiced (t=2.30, p=0.022) and proficiency (Intermediate: t=-2.50, p=0.019; Advanced: t=-4.12, p<0.001) had a significant effect on initial performance. Once again, participants with higher proficiency had better initial performance. No factors had a significant effect on learning rate for the total cost. Overall, the effects of our total cost metric were simply an average of the match and time cost metrics and did not have any unique insights. As such, looking at sequential accuracy and rhythmic accuracy independently may be the most meaningful approach for evaluations in future studies of PHL. This result is expected; previous PHL studies showed stronger and more immediate effects on match cost than time cost.

The full list of significance tests, including the 95% confidence intervals and the estimated effect sizes for the learning curve are listed in Table 1. Considering the significant effects of proficiency and condition on learning rate with the matching cost, we were interested in further evaluating how the learning pattern may be affected by the factors. Particularly, we compared the estimated levels of initial performance and learning rate depending on proficiency and condition using estimated marginal means for their main effect levels and their linear trend by day.

	Meaning	*Log. Match Cost*	*Log. Time Cost*	*Log. Total Cost*
Random intercept	Initial cost due to participant	**3.94***, p<0.001 [3.22, 4.67]**	**3.09***, p<0.001 [2.46, 3.73]**	**4.32***, p<0.001 [3.58, 5.06]**
Condition [stimulus]	Initial cost due to condition	0.29+, p=0.086 [-0.04, 0.62]	0.04, p=0.778 [-0.23, 0.31]	0.20, p=0.201 [-0.11, 0.52]
Music [SYMD]	Initial cost due to music	0.28+, p=0.100 [-0.05, 0.61]	**0.40**, p=0.004 [0.13, 0.67]**	**0.37*, p=0.022 [0.05, 0.68]**
Block	Initial cost due to block	**-0.34*, p=0.049 [-0.67, 0.00]**	0.14, p=0.332 [-0.14, 0.41]	-0.14, p=0.388 [-0.46, 0.18]
Proficiency [Intermediate]	Initial cost due to intermediate proficiency	**-1.10*, p=0.013 [-1.96, -0.23]**	**-1.09**, p=0.008 [-1.88, -0.29]**	**-1.09*, p=0.020 [-2.01, -0.17]**
Proficiency [Advanced]	Initial cost due to advanced proficiency	**-1.64***, p<0.001 [-2.42, -0.86]**	**-1.67***, p<0.001 [-2.39, -0.95]**	**-1.61***, p<0.001 [-2.44, -0.78]**
Day	Overall learning rate per day	**-0.29***, p<0.001 [-0.41, -0.16]**	**-0.28***, p<0.001 [-0.39, -0.18]**	**-0.31***, p<0.001 [-0.43, -0.19]**
Day × Condition [stimulus]	Learning rate due to condition	**-0.09*, p=0.013 [-0.17, -0.02]**	0.00, p=0.935 [-0.06, 0.06]	-0.06+, p=0.080 [-0.13, 0.01]
Day × Music [SYMD]	Learning rate due to music	-0.04, p=0.288 [-0.11, 0.03]	0.00, p=0.921 [-0.06, 0.06]	-0.02, p=0.648 [0.09, 0.05]
Day × Block	Learning rate due to block	0.04, p=0.291 [-0.03, 0.12]	-0.03, p=0.346 [-0.09, 0.03]	0.01, p=0.743 [-0.06, 0.08]
Day × Proficiency [Intermediate]	Learning rate due to intermediate proficiency	0.06, p=0.191 [-0.03, 0.15]	**0.13***, p<0.001 [0.05, 0.21]**	0.07, p=0.105 [-0.02, 0.16]
Day × Proficiency [Advanced]	Learning rate due to advanced proficiency	**0.11**, p=0.007 [0.03, 0.20]**	**0.12***, p<0.001 [0.05, 0.19]**	0.07+, p=0.070 [-0.01, 0.15]

Table 1. Estimates, p-values and 95% confidence intervals for each effect being estimated by linear mixed-effects models (LMEs) for the learning curve. Separate LMEs were constructed for logarithmic match, time and total cost. First six terms relate to overall or initial performance, while later five terms relate to learning rate. Significance is indicated by the p-values and bold entries, marked as p<0.001 (***), p<0.01 (**), p<0.05 (*), p<0.1 (+).

Proficiency	Condition	Initial Performance	Daily Learning Rate	Simplified Equation	Days to Mastery
Beginner	Control	3.33 ± 0.268 (27.94)	0.248 ± 0.0347	$y = e^{3.33-0.248t}$	10.63
	Stimulus	3.59 ± 0.268 (36.23)	0.342 ± 0.0341	$y = e^{3.59-0.342t}$	8.47
Intermediate	Control	2.29 ± 0.338 (9.87)	0.186 ± 0.0420	$y = e^{2.29-0.186t}$	8.59
	Stimulus	2.49 ± 0.337 (12.06)	0.280 ± 0.0413	$y = e^{2.49-0.280t}$	6.42
Advanced	Control	1.81 ± 0.286 (6.11)	0.133 ± 0.0361	$y = e^{1.81-0.133t}$	8.40
	Stimulus	2.00 ± 0.287 (7.39)	0.227 ± 0.0364	$y = e^{2.00-0.227t}$	5.76

Table 2. Initial performance, learning rate and days required to reach mastery (<2% matching error) for each proficiency level and condition. The initial performance and daily learning rate are shown with standard error. The initial performance additionally has the linear equivalent for the log-scale term in parentheses.

While the effect of condition was not significant on initial performance, the estimated initial performance by condition was 2.48 ± 0.181 during control and 2.67 ± 0.181 during stimulus. This result was likely a non-significant artifact of the imbalance between factors. For proficiency, beginners had an estimated initial performance of 3.43 ± 0.260, intermediate participants had an estimated initial performance of 2.39 ± 0.330 and advanced participants had an estimated initial performance of 1.90 ± 0.279. We converted the results to the normalized linear matching cost, finding that beginners started with 30.9% matching cost, intermediate participants started with 10.9% and advanced participants started with 6.7%. Proficiency was a very important factor in predicting performance, and greatly altered how many mistakes participants made even when they had just started learning a new song.

To determine learning rate, we inverted the sign of the estimated effect measured by the model as a negative effect by day indicated a positive learning rate (i.e., making less mistakes over time). The estimated learning rate of participants during the control condition was 0.189 ± 0.0268 whereas during the stimulus condition it was 0.283 ± 0.0263. Comparatively, the stimulus condition where participants received PHR had a learning rate that was greater by **49.7%**, indicating the powerful impact that PHR has on participants' learning rate. Proficiency also remained an important predictor of learning, where beginners had a learning rate of 0.295 ± 0.0289, intermediate participants had a learning rate of 0.233 ± 0.0372 and advanced participants had a learning rate of 0.180 ± 0.0310. More proficient participants' were able to correct errors less quickly, but reached mastery on a new song earlier than their peers.

We also calculated the estimates for each combination of proficiency and condition. These estimates could be used to build a simplified version of the equation used by the LME that ignores random effects and averages additional factors, as shown in Table 2. The simplified equation can be used to approximate how many days it takes to reach mastery for the average participant of a given proficiency with and without PHR, based on when they would reach a normalized matching cost of 2%. As seen in the table, participants across proficiency levels were able to benefit from PHR and had a faster learning rate in PHR. Moreover, participants across all levels would take around two days less to reach mastery on the songs practiced when using PHR. Thus, PHR could be used to practically accelerate learning by reducing the amount of time it takes to reach mastery when learning a new song.

5.2 Improvement Metrics

After developing our learning curve, we also designed LMEs for evaluating particular aspects of the effect of PHR. For these models, we used differential metrics of improvement (described in Section 4.3) between specific recordings as used by past studies [8, 47]. These models only tested the main effect of each factor (including days of practice) as they were differential scores and did not have the time-based component of learning rate. The measures that were compared are shown visually in Figure 7.

5.2.1 Passive Rehearsal and Active Practice. To answer RQ2, we tested the independent effects of passive rehearsal and active practice. We were interested in the effect of PHR and the recall between sessions with and without PHR, so we compared their cost before a new day of active practice to their cost on the previous day. For match cost, the effect of condition (control: M=-0.12%, SE=0.691, stimulus: M=1.87%, SE=1.08) was significant (t=2.73, p=0.007). For time cost, the effect of condition (control: M=-0.37%, SE=0.325, stimulus: M=-0.45%, SE=0.418) was not significant (t=0.35, p=0.73). Finally, for total cost, the effect of condition (control: M=-0.55%, SE=0.89, stimulus: M=1.42%, SE=1.13) was significant (t=2.62, p=0.009). The most interesting component of this result was the inversion in direction for the difference in match and total cost – participants' forgetting effects were negated and they continued to learn between their practice sessions. We tested whether this effect was significant by comparing the stimulus condition to the null hypothesis with a one-sample t-test (i.e., we compare the amount of learning improvement to 0 to show that learning - not just reducing forgetting - is occurring during passive practice). We found that for match cost, stimulus was significantly different from no change (t=2.05, p=0.049). For total cost, stimulus was not significantly different from no change (t=1.35, p=0.19). Based on matching cost, PHR significantly improved performance between sessions, going beyond eliminating recall issues. This result is the most promising sign for the two ways PHR could help with piano education: accelerating learning and encouraging better recall.

For active practice, we were curious to see if the effectiveness of active practice would be impacted by the use of PHR, so we measured the effect immediately before and after an active practice session. For match cost, the effect of condition (control: M=6.34%, SE=1.00, stimulus: M=5.55%, SE=0.99) was not significant (t=-1.39, p=0.32). For time cost, the effect of condition (control: M=2.96%, SE=0.396, stimulus: M=2.98%, SE=0.413) was not significant (t=-0.49, p=0.62) either. Likewise, the effect of condition (control: M=9.74%, SE=1.07, stimulus: M=8.54%, SE=1.10) on total cost was not significant (t=-1.53, p=0.13). Overall, we do not see an effect of PHL on the effectiveness of active practice.

5.2.2 Daily Practice and Longitudinal Improvement. For comparability with past studies, we evaluated the effect of PHR within a day and over the entire week of practice. To measure the combined effect of PHR on daily practice, we evaluated the difference between post-practice scores. For the match cost, we found that the effect of condition (control: M=2.77%, SE=0.606; stimulus: M=4.11%, SE=0.869) was marginally insignificant (t=1.87, p=0.06). For the time cost, the effect of condition (control: M=2.21%, SE=0.356; stimulus: M=2.20%, SE=0.394) was not significant (t=0.63, p=0.53). Finally, for total cost, we found that the effect of condition (control: M=5.18%, SE=0.729; stimulus: M=6.32%, SE=1.00) was stronger than time but also not significant (t=1.73, p=0.08). Overall,

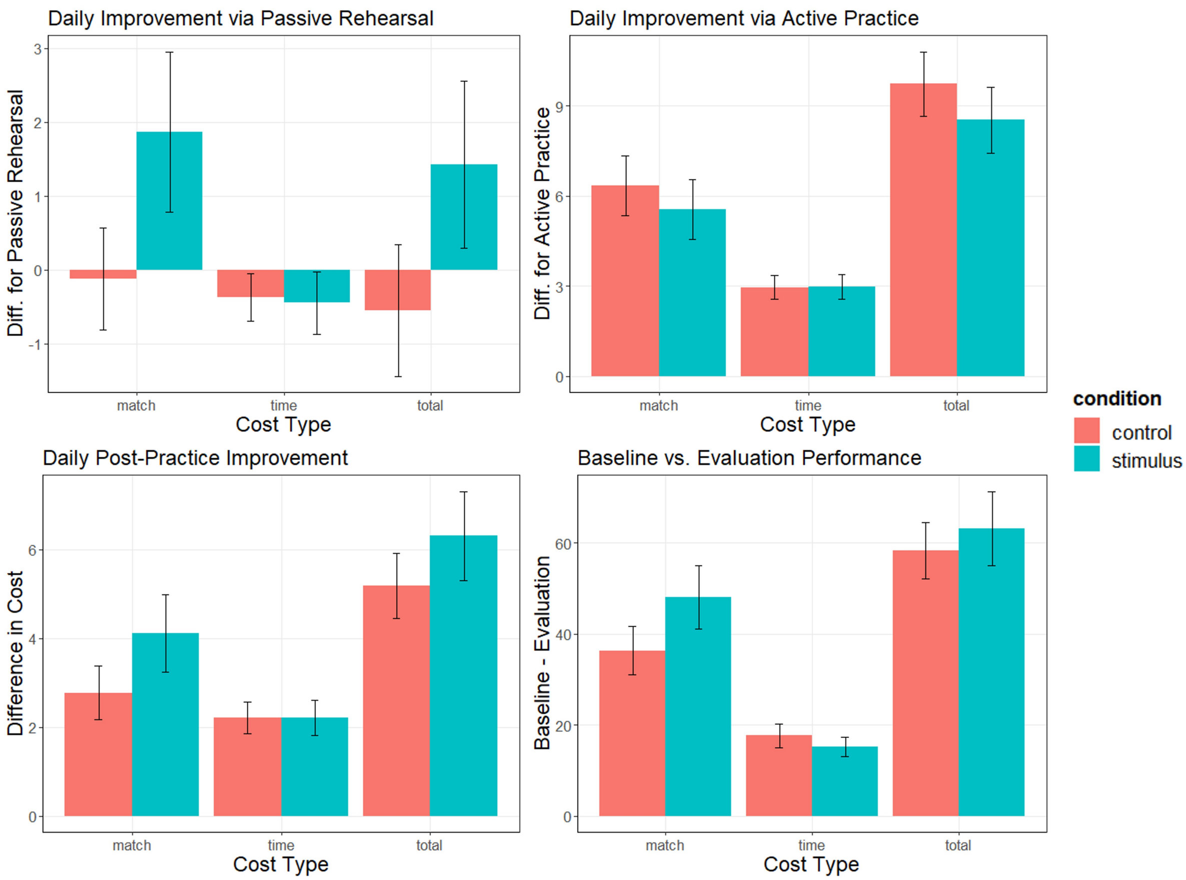

Fig. 7. Improvement of participants' performance. Top left: from left to right showing: Difference between each day of practice, difference between evaluation and baseline, difference between performance before and after active practice, and difference before and after daily gaps in active practice.

the match cost results were somewhat weaker than the learning curve because LMEs are most effective when modeling raw, normally distributed data. However, the means except time cost still showed a large difference in cost improvement per session with PHR.

To measure the longitudinal effect of PHR on participants' improvement over the study duration, we compared their baseline to their evaluation results. We note that this metric had less statistical power due to fewer samples compared to the daily metrics, as it was once per-participant. For match cost, we found that the effect of condition (control: M=36.34%, SE=5.29, stimulus: M=63.16%, SE=6.98) was not significant (t=1.92, p=0.07). However, for the time cost, we found that the effect of condition (control: M=17.64%, SE=2.63, stimulus: M=15.11%, SE=2.12) was significant but in the opposite direction (t=-2.40, p=0.03). As for the total cost, we found that the effect of condition (control: M=58.37%, SE=6.17, stimulus: M=63.16%, SE=8.18) was not significant (t=0.49, p=0.64). The effect on time cost in final performance was a surprising result that was likely created due to how time cost is calculated in our system. As participants practiced, their mistakes stopped being registered as matching mistakes, allowing the relative timing to take priority since they were at least in the correct sequence.

5.3 Qualitative Results

Based on interviews conducted halfway through and upon the conclusion of each participant's study, we obtained qualitative feedback on the study and the PHL gloves. Feedback was generally very positive. Participants enjoyed the study, found the technology exciting, and only had minor complaints about the systems. One participant stated "I think the study was great and I want to play piano more – so that's cool. It showed me that I can actually improve on something that I practice" (P8). When reflecting on the study, most participants felt that the gloves helped them learn the piece faster. One participant said "I think the gloves helped me learn faster…I feel like it did because I'm very surprised how quickly I learned it" (P30). Some participants were unsure if the gloves had improved their performance legitimately or as a placebo: "I don't know if it's placebo or not, but, I think on the first day, I felt that it might have been slightly easier [with the gloves]" (P14). A small number of participants said the gloves seemed to have no effect at all. There wasn't a clear way for participants to tell if the gloves were accelerating their learning apart from their own self-reflection. Because missed days were common and expected due to the nature of a two-week-long in-the-wild study, we gave participants some leniency as long as delays weren't excessive (i.e. more than three days). As explained in Section 4.1, some participants had to be dismissed from the study due to missed days. We established text message check-ins to ensure participants were on track, answer their questions at any time, and catch problems early. When the PHL gloves became damaged during the study, researchers would need to collect and repair the gloves, and have them returned in less than a day. At a larger scale, these processes would quickly become impractical and necessitate a more structured protocol.

Participants generally reported that the gloves were comfortable enough to wear after getting used to them: "They were actually surprisingly not that bad. First couple of days I was getting used to it. In the middle it was like, okay, I'm just wearing these" (P32). To some, the vibrations felt intense at first: "I was shocked by how much vibration there was, and it prevented me from studying a little bit. But afterwards, I think the buzzing just became normalized" (P8). Participants who felt the vibrations were too intense sometimes broke up PHL sessions into smaller chunks such as two 1-hour sessions or four 30-minute sessions. Some participants mentioned that the fingers may be too loose, and direct contact between their fingers and the motors was not always maintained. Many participants also felt worried about breaking the gloves since the wires were exposed between the microcontroller and gloves: "The gloves were comfortable… Picking stuff up, I was kind of nervous about damaging the gloves" (P11). About one-third of participants ended up inadvertently breaking a wire connection on the gloves due to the gloves' fragility. Some participants didn't notice a wire was broken; instead, the breakage was noticed by a researcher during the follow-up or final lab session.

6 DISCUSSION

6.1 Encouraging Recall or Accelerating Learning

While using PHR, participants were able to invert the forgetting effect and learn between active practice sessions, as we showed in Section 5.2.1. As Donchev et al. have previously reported [8], part of this effect is due to encouraging recall via PHL sessions. The differences we measured for the decay or improvement in performance between active sessions provides some insight into where the effect was due to recall and where it was due to learning. Participants had a minor amount (0.12%) of forgetting in the control condition for matching cost, while they had a much larger improvement (1.87%) during the stimulus condition. The reduced forgetting in our participants compared to the average 1.5 errors with aided recall in Donchev et al.'s study was unsurprising, as our participants had sheet music they could rely on and only had a one-day gap between active practice sessions instead of three days. The positive direction and size of the improvement indicate that for matching cost, participants went far beyond increasing their recall and continued to learn during passive rehearsal sessions.

Due to our study design including daily practice sessions, the benefits of accelerating learning were far more apparent in the participants' practice sessions than the recall benefits. However, for future applications of PHR, the recall benefits may be more valuable, sustaining knowledge when piano learners have less time for active practice due to their busy lifestyles. The learning and recall benefits can also be personalized towards the goals of each learner – for example, participants who are trying to learn a new song would receive stimuli for exclusively that song to accelerate their learning. On the other hand, learners who will not be able to practice for an extended period of time can switch the balance towards recall, receiving stimuli on various songs that they still wish to be able to play when they return to actively practicing piano. Therefore, the learning and recall benefits increase the flexibility of piano practice and may help sustain the motivation of piano learners by adapting to their lifestyle.

6.2 Insights from Designing Haptic Gloves for Daily Wear

Through interviews during the study, we identified three areas of improvement for future glove iterations: usability, modularity, and personalizability.

A prevalent issue was the durability of the gloves, particularly with the wire assembly for the haptic motor connections. The wire assembly was not robust enough to handle daily activities, and led to operational failures that would disrupt the progression of the study. The exposed wire assembly also led to some perception that the gloves could not handle these activities, hampering the goal of PHR in the wild. The electronics casing as well as the synthetic leather material of the glove were also lowlights for usability, due to bulkiness and breathability. Moreover, the integration of hardware into the glove impeded its cleanability, posing a challenge for maintenance and hygiene upkeep.

Some participants found the vibration to be too strong, while others found it relaxing. During pilot testing, we noticed that some vibration patterns made learning difficult, such as when simultaneous notes were spaced to be sequential instead. Despite difficulties in perceiving simultaneous vibrations, it appears to be more effective for learning. The vibration pattern also needs to accurately reflect the fingers pressed during piano practice, which may differ due to hand size and habit. Future PHL gloves must allow users to personalize the vibration strength and pattern according to their preferences and piano habits while ensuring that the pattern still facilitates learning.

The insights gained from the user studies underscore the importance of addressing design flaws to enhance the performance, comfort, and practicality of the PHL glove prototype – especially as we work towards the ultimate goal of the PHL gloves being a viable consumer product. Informed by these findings from the user studies, future work requires a robust, compact, and fully flexible design for the PHL gloves to overcome the previous iteration's shortcomings and support longitudinal studies.

6.3 Advancing Evaluations of Learning in HCI

There is an important distinction between performance and learning. As we've shown, performance in a single musical piece can be depicted by an exponential decay function, where error is reduced over time. In the absence of learning curves, evaluations in past studies of musical learning (including PHL) in HCI generally use three methods: unaided vs. aided performance [28, 39], unaided performance before vs. after session with aid [12, 47] or time to mastery [8, 18, 28]. Given the exponential form, time taken to reach mastery is by far the most meaningful metric, as it captures the essence of exponential decay in a similar manner with half-life. However, time to mastery is challenging for controlled studies due to the variable duration. Unaided performance before and after a session is a meaningful substitute, as it captures the effect, although in a linear manner. We recommend that future studies also take note of how proficiency may affect performance when evaluating learning. On the other hand, we note that directly comparing unaided vs. aided performance gives very little information about learning despite its use in the past, particularly for studies involving active feedback or visualizations. It biases the results with the *performance* while using the aid instead of the *learning* while using the aid. Thus, any performance aids in a musical learning system should be disabled during evaluations as part of the study design. We strongly encourage future researchers of musical interfaces to use the time to mastery metric, or the unaided difference in accuracy and error for their evaluations.

6.4 Limitations

6.4.1 Musical Pieces Were Not Truly Equivalent. Although the two piano pieces, CAP and SYMD, were similar in difficulty, as explained in Section 4.2, participants almost always reported SYMD was harder to learn. This was likely due to the higher number of notes requiring the player to use the sharps or flats on the keyboard and the greater independence of the left and right hand as both were often playing different rhythms. Having one song that was more difficult affected the time cost of participants, and may have had an effect on participants' frustration and learning ability. We were able to balance this by alternating which song participants started with and the experimental condition used on each song, but it may have affected the analysis of individual participants.

6.4.2 Skill Generalizability Remains Unclear. In our study, we mainly evaluated whether the correct notes were pressed at the correct time. However, in the real world, piano mastery requires much more than just these skills; it requires more nuanced skills such as dynamics, expression, phrasing, hand movements, and so on. Furthermore, as the PHR system reinforces fingerings and timings for one specific song at a time, it may not impact the learner's skills on any other song. In other words, PHR may reinforce musical skills for one song, but might not translate to an acquisition of piano skills in general. It is unclear if PHR can be used to teach or even accelerate the learning of general musicianship for pianists, or if the skills gained from learning one song with a PHR approach can benefit the learner on a new piece as well. For teaching generalizable skills to beginners, practicing major scales – as typically done in piano education – could be more effective than practicing a song.

6.4.3 Musical Pieces Used are Too Easy for Advanced Piano Players. As described in Section 4.1, participants who were able to perform the songs perfectly on their first sight-reading were not allowed to continue the study because they had no room to improve. However, some advanced players who were not excluded still achieved mastery on their second or third day of practicing, which limited the amount of data available to analyze. Ideally, advanced players would have more challenging songs to practice, allowing for more robust modeling of the effects of PHR. Based on qualitative feedback during the study, advanced players may prefer using the gloves for recall rather than learning. Therefore, rehearsing sub-sequences of longer songs to encourage recall may be more beneficial than trying to passively teach new songs to expert pianists.

7 CONCLUSION

In this paper, we proposed and demonstrated a new approach to music education combining PHL with deliberate active practice, which we call passive haptic rehearsal (PHR). We conducted the first longitudinal in-the-wild study of PHL, where 36 participants with a musical background practiced two songs for one week each, with and without PHR. To support the study and practical usage of PHL, we developed an online portal and haptic gloves compatible with daily wear. Based on the 20 participants with complete and valid data, we modeled and analyzed the influence of PHR on the learning curve of playing new songs, finding that PHR significantly increased the learning rate for participants' matching accuracy, increasing the learning rate by 49.7% and reducing the days needed to reach mastery. However, we found that PHR did not affect participants' ability to learn rhythm. We confirmed that PHL boosts long-term recall and had no effect on the effectiveness of active practice sessions.

We discussed new innovations for increasing the ease of deploying PHL beyond the laboratory setting. In the future, longitudinal studies, including multiple songs or a musical curriculum, are necessary to evaluate the integration of PHR within musical education. To facilitate these, we are developing a robust, fully flexible glove using a flexPCB and developing a pipeline to automatically generate vibration sequences for all piano sheet music. Finally, to sustain students' motivation to practice, further studies can benefit from gamified progress tracking and personalized vibration patterns. This work brings PHL closer to consumer applications beyond research, situated as part of a flexible and adaptive piano education system.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation, Partnerships for Innovation grant #2122797, and Convergence Accelerator grant #2236014. This research was also supported by the Georgia Research Alliance, Inc. based in Atlanta, Georgia.

We are grateful to Seth Radman for his guidance on musical technology in the industry. We would like to thank Gale Wolfe, Prof. Gil Weinberg, and Prof. Andrea McAlister for their advice on piano pedagogy. Lastly, we would like to thank the many Georgia Institute of Technology students who have contributed to various versions of the passive haptic learning gloves and web portal, particularly Noah Teuscher, Soo Bin Park, Emerson Miller, Diego De Dios Suarez, and Barnabas Li.

REFERENCES

[1] Timothy J. Aveni, Caitlyn Seim, and Thad Starner. 2019. A Preliminary Apparatus and Teaching Structure for Passive Tactile Training of Stenography. In *2019 IEEE World Haptics Conference (WHC)*. 383–388. https://doi.org/10.1109/WHC.2019.8816077

[2] Douglas Bates, Martin Maechler, Ben Bolker, Steven Walker, Rune Haubo Bojesen Christensen, Henrik Singmann, Bin Dai, Gabor Grothendieck, Peter Green, and M Ben Bolker. 2015. Package 'lme4'. *convergence* 12, 1 (2015), 2.

[3] Sissela Bergman Nutley, Fahimeh Darki, and Torkel Klingberg. 2014. Music Practice Is Associated with Development of Working Memory during Childhood and Adolescence. *Frontiers in Human Neuroscience* 7 (Jan. 2014). https://doi.org/10.3389/fnhum.2013.00926

[4] Asha Bhandarkar, Tan Gemicioglu, Brahmi Dwivedi, Caitlyn Seim, and Thad Starner. 2022. Learning Piano Songs with Passive Haptic Training: An Interactive Lesson. In *Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, Cambridge United Kingdom, 10–12. https://doi.org/10.1145/3544793.3560321

[5] Eugenia Costa-Giomi, Patricia J. Flowers, and Wakaha Sasaki. 2005. Piano Lessons of Beginning Students Who Persist or Drop Out: Teacher Behavior, Student Behavior, and Lesson Progress. *Journal of Research in Music Education* 53, 3 (Oct. 2005), 234–247. https://doi.org/10.1177/002242940505300305

[6] Michael Scott Cuthbert and Christopher Ariza. 2010. music21: A toolkit for computer-aided musicology and symbolic music data. (2010).

[7] Jordan Aiko Deja, Sven Mayer, Klen Čopič Pucihar, and Matjaž Kljun. 2022. A Survey of Augmented Piano Prototypes: Has Augmentation Improved Learning Experiences? *Proceedings of the ACM on Human-Computer Interaction* 6, ISS (Nov. 2022), 566:226–566:253. https://doi.org/10.1145/3567719

[8] Rumen Donchev, Erik Pescara, and Michael Beigl. 2021. Investigating Retention in Passive Haptic Learning of Piano Songs. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 2 (June 2021), 60:1–60:14. https://doi.org/10.1145/3463513

[9] Carolyn Drake and Caroline Palmer. 2000. Skill Acquisition in Music Performance: Relations between Planning and Temporal Control. *Cognition* 74, 1 (Jan. 2000), 1–32. https://doi.org/10.1016/S0010-0277(99)00061-X

[10] K. Anders Ericsson, Ralf T. Krampe, and Clemens Tesch-Römer. 1993. The Role of Deliberate Practice in the Acquisition of Expert Performance. *Psychological Review* 100, 3 (1993), 363–406. https://doi.org/10.1037/0033-295X.100.3.363

[11] LTC Tanya Estes, Deborah Backus, and Thad Starner. 2015. A wearable vibration glove for improving hand sensation in persons with Spinal Cord Injury using Passive Haptic Rehabilitation. In *2015 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*. 37–44. https://doi.org/10.4108/icst.pervasivehealth.2015.259137

[12] Likun Fang, Timo Müller, Erik Pescara, Nikola Fischer, Yiran Huang, and Michael Beigl. 2023. Investigating Passive Haptic Learning of Piano Songs Using Three Tactile Sensations of Vibration, Stroking and Tapping. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 3 (Sept. 2023), 95:1–95:19. https://doi.org/10.1145/3610899

[13] Likun Fang, Erik Pescara, Patrick Reiter, and Michael Beigl. 2023. How Could I Learn Rhythm Better? Investigating Three Learning Signals for Passive Haptic Learning in Different Context. In *Proceedings of the 16th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '23)*. Association for Computing Machinery, New York, NY, USA, 279–283. https://doi.org/10.1145/3594806.3594807

[14] Tan Gemicioglu, Noah Teuscher, Brahmi Dwivedi, Soobin Park, Emerson Miller, Celeste Mason, Caitlyn Seim, and Thad Starner. 2022. Passive Haptic Rehearsal for Accelerated Piano Skill Acquisition. https://doi.org/10.48550/arXiv.2203.12749 arXiv:2203.12749 [cs]

[15] Michael Good. 2001. MusicXML for Notation and Analysis. *The virtual score: representation, retrieval, restoration* 12, 113–124 (2001), 160.

[16] A. Heathcote, S. Brown, and D. J. Mewhort. 2000. The Power Law Repealed: The Case for an Exponential Law of Practice. *Psychonomic Bulletin & Review* 7, 2 (June 2000), 185–207. https://doi.org/10.3758/bf03212979

[17] Kevin Huang, Ellen Yi-Luen Do, and Thad Starner. 2008. PianoTouch: A Wearable Haptic Piano Instruction System for Passive Learning of Piano Skills. In *2008 12th IEEE International Symposium on Wearable Computers*. 41–44. https://doi.org/10.1109/ISWC.2008.4911582

[18] Kevin Huang, Thad Starner, Ellen Do, Gil Weiberg, Daniel Kohlsdorf, Claas Ahlrichs, and Ruediger Leibrandt. 2010. Mobile Music Touch: Mobile Tactile Stimulation for Passive Learning. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems - CHI '10*. ACM Press, Atlanta, Georgia, USA, 791. https://doi.org/10.1145/1753326.1753443

[19] Jakob Karolus, Annika Kilian, Thomas Kosch, Albrecht Schmidt, and Paweł W. Wozniak. 2020. Hit the Thumb Jack! Using Electromyography to Augment the Piano Keyboard. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference (DIS '20)*. Association for Computing Machinery, New York, NY, USA, 429–440. https://doi.org/10.1145/3357236.3395500

[20] Joel J. Katz, Momo Ando, and Melody Wiseheart. 2021. Optimizing Song Retention through the Spacing Effect. *Cognitive Research: Principles and Implications* 6, 1 (Dec. 2021), 79. https://doi.org/10.1186/s41235-021-00345-7

[21] Daniel Kohlsdorf and Thad Starner. 2010. Mobile Music Touch: The Effect of Primary Tasks on Passively Learning Piano Sequences. In *International Symposium on Wearable Computers (ISWC) 2010*. 1–8. https://doi.org/10.1109/ISWC.2010.5665877

[22] Herbert E. Krugman and Eugene L. Hartley. 1970. Passive Learning from Television. *Public Opinion Quarterly* 34, 2 (Jan. 1970), 184–190. https://doi.org/10.1086/267788

[23] Alexandra Kuznetsova, Per B. Brockhoff, and Rune H. B. Christensen. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82 (Dec. 2017), 1–26. https://doi.org/10.18637/jss.v082.i13

[24] Nathaniel Leibowitz, Barak Baum, Giora Enden, and Amir Karniel. 2010. The Exponential Learning Equation as a Function of Successful Trials Results in Sigmoid Performance. *Journal of Mathematical Psychology* 54, 3 (June 2010), 338–340. https://doi.org/10.1016/j.jmp.2010.01.006

[25] Granit Luzhnica. 2019. *Leveraging Optimisations on Spatial Acuity for Conveying Information through Wearable Vibrotactile Display*. Ph. D. Dissertation. https://doi.org/10.13140/RG.2.2.34648.24320

[26] Granit Luzhnica and Eduardo Veas. 2019. Optimising Encoding for Vibrotactile Skin Reading. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300465

[27] Tanya Thais Markow. 2012. Mobile Music Touch: Using Haptic Stimulation for Passive Rehabilitation and Learning. (March 2012).

[28] Karola Marky, Andreas Weiß, Andrii Matviienko, Florian Brandherm, Sebastian Wolf, Martin Schmitz, Florian Krell, Florian Müller, Max Mühlhäuser, and Thomas Kosch. 2021. Let's Frets! Assisting Guitar Students During Practice via Capacitive Sensing. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3411764.3445595

[29] Tríona McCaffrey and Jane Edwards. 2016. "Music Therapy Helped Me Get Back Doing": Perspectives of Music Therapy Participants in Mental Health Services. *Journal of Music Therapy* 53, 2 (2016), 121–148. https://doi.org/10.1093/jmt/thw002

[30] Marco Musy. 2022. PianoPlayer.

[31] National Endowment for the Arts, United States. Bureau of the Census, and United States. Bureau of Labor Statistics. 2021. Arts Basic Survey, United States, 2020. https://doi.org/10.3886/ICPSR37972.v1

[32] Saul B. Needleman and Christian D. Wunsch. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of molecular biology* 48 3 (1970), 443–53.

[33] Arinobu Niijima. 2023. Learning Effects and Retention of Electrical Muscle Stimulation in Piano Playing. In *Proceedings of the 2023 ACM International Symposium on Wearable Computers (ISWC '23)*. Association for Computing Machinery, New York, NY, USA, 104–108. https://doi.org/10.1145/3594738.3611373

[34] Jun Nishida, Yudai Tanaka, Romain Nith, and Pedro Lopes. 2022. DigituSync: A Dual-User Passive Exoskeleton Glove That Adaptively Shares Hand Gestures. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology (UIST '22)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3526113.3545630

[35] Se-Woong Park, Tjeerd Dijkstra, and Dagmar Sternad. 2013. Learning to Never Forget—Time Scales and Specificity of Long-Term Memory of a Motor Skill. *Frontiers in Computational Neuroscience* 7 (Sept. 2013). https://doi.org/10.3389/fncom.2013.00111

[36] R Core Team. 2023. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

[37] Linsey Raymaekers, Jo Vermeulen, Kris Luyten, and Karin Coninx. 2014. Game of Tones: Learning to Play Songs on a Piano Using Projected Instructions and Games. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems (CHI EA '14)*. Association for Computing Machinery, New York, NY, USA, 411–414. https://doi.org/10.1145/2559206.2574799

[38] Henry L. Roediger and Megan A. Smith. 2012. The "Pure-Study" Learning Curve: The Learning Curve without Cumulative Testing. *Memory & Cognition* 40, 7 (Oct. 2012), 989–1002. https://doi.org/10.3758/s13421-012-0213-5

[39] Katja Rogers, Amrei Röhlig, Matthias Weing, Jan Gugenheimer, Bastian Könings, Melina Klepsch, Florian Schaub, Enrico Rukzio, Tina Seufert, and Michael Weber. 2014. P.I.A.N.O.: Faster Piano Learning with Interactive Projection. In *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces (ITS '14)*. Association for Computing Machinery, New York, NY, USA, 149–158. https://doi.org/10.1145/2669485.2669514

[40] Kathryn Roulston, Peter Jutras, and Seon Joo Kim. 2015. Adult Perspectives of Learning Musical Instruments. *International Journal of Music Education* 33, 3 (Aug. 2015), 325–335. https://doi.org/10.1177/0255761415584291

[41] Nicolas Ruth and Daniel Müllensiefen. 2021. Survival of Musical Activities. When Do Young People Stop Making Music? *PLoS ONE* 16, 11 (Nov. 2021), e0259105. https://doi.org/10.1371/journal.pone.0259105

[42] Caitlyn Seim. 2019. *Wearable Vibrotactile Stimulation: How Passive Stimulation Can Train and Rehabilitate.* Ph. D. Dissertation. Georgia Institute of Technology.

[43] Caitlyn Seim, John Chandler, Kayla DesPortes, Siddharth Dhingra, Miru Park, and Thad Starner. 2014. Passive Haptic Learning of Braille Typing. In *Proceedings of the 2014 ACM International Symposium on Wearable Computers (ISWC '14)*. Association for Computing Machinery, New York, NY, USA, 111–118. https://doi.org/10.1145/2634317.2634330

[44] Caitlyn Seim, Bingxian Chen, Chuzhang Han, David Vacek, Alexis Lowber, Maarten Lansberg, and Allison M. Okamura. 2023. Daily Vibrotactile Stimulation Exhibits Equal or Greater Spasticity Relief Than Botulinum Toxin in Stroke. *Archives of Physical Medicine and Rehabilitation* 104, 10 (Oct. 2023), 1565–1572. https://doi.org/10.1016/j.apmr.2023.03.031

[45] Caitlyn Seim, Bingxian Chen, Chuzhang Han, David Vacek, Laura Song Wu, Maarten Lansberg, and Allison Okamura. 2023. Relief of Post-Stroke Spasticity with Acute Vibrotactile Stimulation: Controlled Crossover Study of Muscle and Skin Stimulus Methods. *Frontiers in Human Neuroscience* 17 (Aug. 2023). https://doi.org/10.3389/fnhum.2023.1206027

[46] Caitlyn Seim, Nick Doering, Yang Zhang, Wolfgang Stuerzlinger, and Thad Starner. 2017. Passive Haptic Training to Improve Speed and Performance on a Keypad. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (Sept. 2017), 100:1–100:13. https://doi.org/10.1145/3132026

[47] Caitlyn Seim, Tanya Estes, and Thad Starner. 2015. Towards Passive Haptic Learning of Piano Songs. In *2015 IEEE World Haptics Conference (WHC)*. 445–450. https://doi.org/10.1109/WHC.2015.7177752

[48] Caitlyn Seim, Rodrigo Pontes, Sanjana Kadiveti, Zaeem Adamjee, Annette Cochran, Timothy Aveni, Peter Presti, and Thad Starner. 2018. Towards Haptic Learning on a Smartwatch. In *Proceedings of the 2018 ACM International Symposium on Wearable Computers (ISWC '18)*. Association for Computing Machinery, New York, NY, USA, 228–229. https://doi.org/10.1145/3267242.3267269

[49] Caitlyn Seim, Saul Reynolds-Haertle, Sarthak Srinivas, and Thad Starner. 2016. Tactile Taps Teach Rhythmic Text Entry: Passive Haptic Learning of Morse Code. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers (ISWC '16)*. Association for Computing Machinery, New York, NY, USA, 164–171. https://doi.org/10.1145/2971763.2971768

[50] Caitlyn E. Seim, David Quigley, and Thad E. Starner. 2014. Passive Haptic Learning of Typing Skills Facilitated by Wearable Computers. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems (CHI EA '14)*. Association for Computing Machinery, New York, NY, USA, 2203–2208. https://doi.org/10.1145/2559206.2581329

[51] Caitlyn E Seim, Brandon Ritter, Thad E Starner, Kara Flavin, Maarten G Lansberg, and Allison M Okamura. 2020. Perspectives on the design and performance of upper-limb wearable stimulation devices for stroke survivors with hemiplegia and spasticity. *bioRxiv* (2020), 2020–08.

[52] John A. Sloboda, Jane W. Davidson, Michael J. A. Howe, and Derek G. Moore. 1996. The Role of Practice in the Development of Performing Musicians. *British Journal of Psychology* 87, 2 (1996), 287–309. https://doi.org/10.1111/j.2044-8295.1996.tb02591.x

[53] Yoshinari Takegawa, Tsutomu Terada, and Masahiko Tsukamoto. 2012. A Piano Learning Support System Considering Rhythm. In *Non-Cochlear Sound: Proceedings of the 38th International Computer Music Conference, ICMC 2012, Ljubljana, Slovenia, September 9-14, 2012*. Michigan Publishing.

[54] Janet van der Linden, Erwin Schoonderwaldt, Jon Bird, and Rose Johnson. 2011. MusicJacket—Combining Motion Capture and Vibrotactile Feedback to Teach Violin Bowing. *IEEE Transactions on Instrumentation and Measurement* 60, 1 (Jan. 2011), 104–113. https://doi.org/10.1109/TIM.2010.2065770

[55] Scott I. Vrieze. 2012. Model Selection and Psychological Theory: A Discussion of the Differences between the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). *Psychological Methods* 17, 2 (2012), 228–243. https://doi.org/10.1037/a0027127

[56] Xiao Xiao and Hiroshi Ishii. 2010. MirrorFugue: Communicating Hand Gesture in Remote Piano Collaboration. In *Proceedings of the Fifth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '11)*. Association for Computing Machinery, New York, NY, USA, 13–20. https://doi.org/10.1145/1935701.1935705

[57] Cliff Zukin and Robin Snyder. 1984. Passive Learning: When the Media Environment Is the Message. *Public Opinion Quarterly* 48, 3 (Jan. 1984), 629–638. https://doi.org/10.1086/268864

A PARTICIPANT BACKGROUND

Subject	Proficiency	Piano Experience	Other Musical Experience
8	Beginner	No	3 years (xylophone, drums, symbols, triangle)
9	Advanced	10 years	No
10	Intermediate	1 years	8 years (clarinet)
11	Advanced	3 years	6 years (guitar)
14	Intermediate	3 years	11 years (cello, saxophone)
16	Advanced	1 years	2 years (bass)
20	Beginner	No	8 years (double bass)
21	Advanced	3 years	6 years (violin)
22	Beginner	No	6 years (cello)
24	Beginner	1 years	3 years (alto saxophone)
25	Advanced	4 years	12 years (cello)
26	Intermediate	No	7 years (trumpet, tuba)
27	Beginner	No	6 years (guitar, bass, ukulele)
28	Beginner	1 years	1 year (recorder)
30	Intermediate	1 years	10 years (cello)
31	Beginner	No	1 year (guitar)
32	Beginner	No	2 years (violin)
33	Intermediate	3 years	No
34	Advanced	4 years	No
35	Advanced	4 years	No

Table 3. Participants and their musical experience.

B SHEET MUSIC

Fig. 8. Sheet music for the two songs used during the study: Béla Bartók's Children at Play (CAP) and Slovak Young Men's Dance (SYMD).

C INDIVIDUAL PARTICIPANT PERFORMANCE

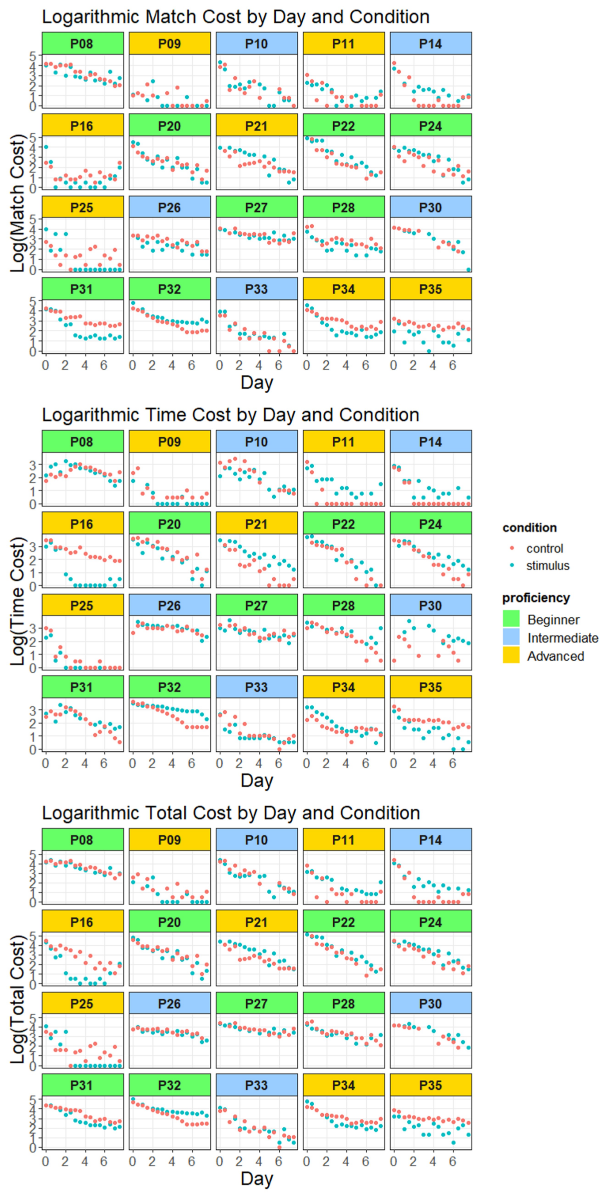

Fig. 9. Logarithmic match, time and total costs over time for each participant. Lower cost indicates better performance, and participants with advanced proficiency often reach mastery early by reducing mistakes to zero. Meanwhile, beginners may never achieve a perfect score. Performance has high variance, and participants sometimes perform worse despite more practice.