Passive Haptic Learning for Vibrotactile Skin-Reading: Comparison of Teaching Structures

Granit Luzhnica
Know-Center
Graz, Austria
gluzhnica@know-center.at

Eduardo Veas
Know-Center
Graz, Austria
eveas@know-center.at

Cailtyn Seim
Georgia Ins't of Technology
Atlanta, GA, USA
ceseim@gatech.edu

ABSTRACT
This paper investigates the effects of using passive haptic learning to train the skill of reading text from vibrotactile patterns. The vibrotactile method of transmitting messages, skin-reading, is effective at conveying rich information but its active training method requires full user attention, is demanding, time-consuming, and tedious. Passive haptic learning offers the possibility to learn in the background while performing another primary task. We present a study investigating the use of passive haptic learning to train for skin-reading. Additionally, a word-based learning structure is typically used for this passive learning method. We expose trends that suggest this word-based incrimental teaching may not be optimal.

ACM Classification Keywords
H.5.2 Information Interfaces and Presentation: User Interfaces—*Haptic I/O*

Author Keywords
passive haptic; haptic feedback; tactile feedback; skin reading; stimulation; haptic display; wearable; user study; HCI

INTRODUCTION
Wearable and mobile devices are already a part of our everyday life. They provide assistance to daily activities and enrich them with additional information collected by the sensors within them. The primary feedback modalities of mobiles and wearables are visual and auditory. Although, most of them contain vibrotactile capabilities, the primary utilisation of vibrotactile feedback is to provide additional support to visual interaction . On the contrary, the vibrotactile feedback on wearable devices has a lot of potential to be used on its own. A prominent application and the subject of this paper is the so called skin reading [12]. *Vibrotactile skin reading* (*VSR*) uses vibrotactile patterns to encode symbols [4, 12, 9, 29] which then can be combined to convey complex messages such as words and phrases [4, 12]. While traditionally, perceiving information through skin has been applied for visually impaired users (e.g Braille reading), users with normal vision can benefit from a means to perceive messages that does not recruit the visual or auditory senses. Typically, learning to associate meanings (e.g letters or words) involves active training, where users receive vibrotactile patterns accompanied by visual and audio cues representing the meaning [12]. Despite being effective, such a training requires full attention, is repetitive and extensive. It is an effort of hours which might pose a motivational obstacle for some non-impaired users.

On the other hand, *passive haptic learning* [15, 17] (*PHL*) can be used to train users passively without requiring their attention. This haptics-based teaching technique has been successfully used in numerous applications such as teaching people to play piano [17] or type in braille [15] without them being actively focused on training. During the training, they are exposed to audio and haptic stimuli that inform a skill, but they need not pay attention to it. Such a technique would be beneficial for training skin-reading as it might motivate potential users of skin reading that are interested but do not have the inclination to go through hours of active training. However, aforementioned studies of PHL are used to train muscle memory, whereas skin-reading is mostly a task of associating a meaning with a tactile pattern. Patterns are combinations of activations of several vibromotors, where failing to perceive one of them may change the meaning of the message. While the encoding method can be optimised to minimise such situations [10], skin-reading requires an amount of concentration in training that may render PHL unusable.

As training is an extensive task, it would be useful if users could be trained with a default transmission speed but be able to understand messages with different transmission speed. This way users could increase the speed over time and they would not need re-training if the speed need to be changed. Perhapse PHL could enable this.

All prior work on PHL for text system (Braille, Morse code, Stenography) learning taught the system incrementally; teaching letters in small groups based on words from a pangram [15, 17]. It was assumed that the small groups and their semantic associations were necessary for learning of many letters; however, no prior work has contrasted this with a training method not requiring semantic grouping. On the other hand, a passive instruction method without having to develop word-based lessons may allow different learning durations, less rigid passive learning structures and less system development. Such non word-based training method without any semantic grouping has been successfully applied in active training for skin reading [12, 11]. In this paper, we contrast these learning structures for PHL.

This paper presents a user study investigating the following research questions:
RQ1: *Can passive haptic learning can be used to train users for skin reading*?
RQ2: *How does the duration of training stimuli affect recognition? Can users understand transmissions at a different speeds than the one used for training*?
RQ3: *Is it necessary to semantically group letters when using PHL as a training method for vibrotactile skin reading? Or would a non word-based training be sufficient*?

RELATED WORK

Starting with Braille's invention of the Braille coding in 1824, tactile displays have been widely used by people with visual impairments. Research on tactile displays equipped with actuators has been ongoing since at least 1924 [3], where Gault [3] used a piezoelectric unit to convert entire recorded speech to touch. Similarly, Kirman [7] used a 15×15 vibrator matrix on the palm to teach six participants to differentiate between the patterns of 15 different words. Other researchers attempted to utilise a visually oriented approach, where a low-resolution image of the object is projected to an array of stimulators. For instance, White [27] transformed images captured from a video feed to a 20×20 vibrotactile display placed on the back. After training, participants were able to distinguish simple shapes like circle, square and triangle. Bliss [1] developed the first commercial device capable of capturing text from the video feed and then imprinting each letter on the finger with a 6×24 matrix of vibrators.

A more successful approach of transmitting information through haptics was provided by Geldard [4] in 1967. The device was named Vibratese and used five vibromotors placed on the chest to encode 45 symbols (letters, numbers and most frequent short words). The author reported that after 65 hours of training one participant was able to understand 38 wpm (words per minute). More recently, Luzhnica et al followed a different encoding scheme using only the location of vibromotors to encode 26 letters of English alphabet [12]. The authors used six vibromotors on the back of the hand and were able to train users to perceive letters, words and phrases within only five hours, although they needed repetition of stimuli.

Information encoding is an important aspect of tactile displays as patterns should be optimised for both discrimination and transmission speed. Typically a combination of variations in amplitude [23, 24, 28], frequency [23, 24, 28], duration [5, 4] and body locations [4, 28, 13, 20] have been used. For instance, Geldard [4] in his Vibratese used five locations, a variation of three durations and three intensities to encode the desired symbols. Recently, Novich [14] showed that spatiotemporal encoding, where vibromotors in a pattern are turned on and off sequentially one after the other, results in significantly better discrimination than the spatially encoded patterns where all vibromotors in a pattern onset simultaneously. Similar findings have been produced by [15, 18]. Liao [9] utilised such a spatiotemporal encoding to encode symbols on the wrist. Although such encoding works well [9, 14] in terms of being identified by participants, it is many times slower than the spatial encoding. Luzhnica [12, 10] used a prioritised overlapping spatiotemporal encoding where vibromotors are activated in sequence after a gap, and they stay on until the pattern is finished. This method resulted in better recognition accuracy than spatial encoding, and it is faster than spatiotemporal encoding, as vibromotors share most of the activated time.

On the other hand, passive haptic learning (PHL) began with simple music sequence training for one hand and has since been explored for multi-limb skills, simultaneous actions, rhythm, other areas of the body and alphabetic codes for text entry [15, 17, 19, 16]. The technique has been found in a limited number of cases and would benefit from further study. This work aims to replicate the technique of PHL and examine it for training users in vibrotactile skin reading. Furthermore, prior work [17] contrasted two teaching structures for passive learning, but this work focused on teaching two-limb skills, and it has not been established whether a semantic chunking structure is beneficial to learning.In practice, the motors could be placed within two wearable sleeves to provide a consumable product.

USER STUDY

The goals of this study were to investigate if PHL is able to train for SkinReading (RQ1), establish the effects of training stimuli speed on recognition results (RQ2), and compare a bottom up, letter by letter training (ABT) with a training based on words cues (WBT, RQ3). We conducted a user study to test reception and knowledge before and after passive training.PHL requires the attention of the user on a primary task while the training takes place in the background. We intend to maintain the study time and attention of participants within manageable margins so, the study uses only ten letters, enough to compose words, while limiting training time to around 30m. Note that, the method used to encode information is not limited to ten letters. It was used by Luzhnica et al. to encode the entire English alphabet, which participants learnt within three hours of active training [12]. This study uses letters: 'A', 'C', 'E', 'G', 'H', 'I', 'M', 'N', 'S' and 'T', encoded with max. two vibromotors (see Figure 1). Given the native German language of the location where user study took place, German was used throughout the study for words and spelling. This study uses two training protocols (RQ3: ABT, WBT), three stimulation speeds during testing (RQ2: 100, 200, 300 ms) and measures of accuracy, repetition of stimuli and testing duration (RQ1).

Wearable Haptic Display Design

We replicate the glove design by Luzhnica et al [12] with six vibromotors on the back of the hand (see Figure 1). With it, the ten letters in the study can be encoded with combinaitons of one or two vibromotors. But, for encoding the entire alphabet, a layout with more vibromotors as proposed by [10] would be a better choice. Our device uses an Arduino Due board to controls $3.4mm$ vibrotactile motors of type ROB-08449 (Voltage range: $2.3V \sim 3.6V$; Amplitude vibration: $0.8G$).

Vibrotactile Patterns and Encoding

Each letter is encoded with one or two vibromotors using an OST (overlapped spatiotemporal) stimulation pattern described by Luzhnica and Veas. [10]. Figure 2 illustrates the

Figure 1: The wearable vibrotactile display layout [12] and the encoding scheme of each letter used during the study.

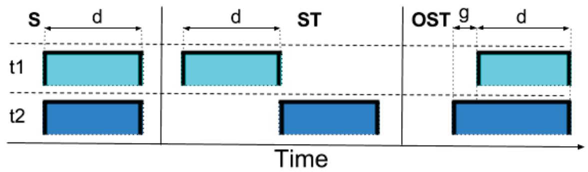


Figure 2: Pattern types composed of two vibromotors/locations: sequential spatiotemporal (ABT), overlapping spatiotemporal (OST), spatial (S). Base duration (d) represents the activation time of a vibromotor (t1 or t2). The gap between the activation of vibromotors is denoted by g.

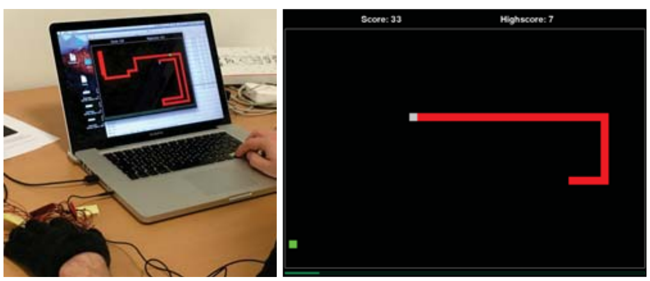


Figure 3: A participant (left) playing the game (right) while being trained to recognise letters using PHL.

details of OST and differences with pure spatial and spatiotemporal stimulation. OST activates the vibromotors in sequence, but they share most of the activation time. Moreover, the order of activation is prioritised by the sensitivity of the finger, since it yields a higher accuracy in identification of locus [10]. Sensitivity order is assumed according to studies suggesting that sensitivity decreases from the index finger towards the little finger: the index finger is more sensitive than the middle, ring, and pinky finger [2, 25, 6]. The thumb is the lowest sensitive [21]. For example, a letter encoded with index and pinky finger, activates the index vibromotor first, and then after a gap, the vibromotor on the pinky finger. Letter encoding uses a base duration (d) of 200 ms and a 10 ms gap (g) between the activation of vibromotors. So, the letter duration (ld) of a one vibromotor letter is 200 ms and 210 ms for two-vibromotor letters. When constructing words, a gap (bl) of 200 ms separates subsequent letters . Note that with longer training periods, users learn to recognise letters and words with shorter stimulation [12]. Our study fixes training duration to 200ms and considers shorter durations during testing.

Procedure

The entire study was organised in *rounds*, each serving the purpose of either training or testing. PHL takes place during training. During the entire training, participants are engaged in playing a game as a primary task (internal implementation of the snake game [1]). Meanwhile, they are passively trained to recognise patterns (see Figure 3). Testing rounds use the active concentration of participant on the test. There are two training modes: WBT and ABT.

Word Based Training (WBT) uses word cues to passively train users to associate letters with vibrotactile patterns. WBT starts with an audio cue of a word (e.g. Ich) and continues with a series of audio cues of each letter of that word. the vibrotactile stimulation of the pattern representing the letter follows 50ms after its audio cue. We use the words "ICH", "MAG", "ES", "NICHT", which together form a sentence ("Ich mag es nicht") from the children's book "Grunes Ei mit Schpeck" written by Dr Seuss. Each word is played in a loop 48 times before moving to the next one. WBT takes 32 minutes.

Alphabetical Based Training (ABT) uses letters in alphabetic order to passively train users to associate symbols with vibrotactile patterns. ABT starts with an audio cue which represents a letter of the German alphabet, followed by its vibrotactile cue after 50 ms. The process is repeated four times, moving to the next letter to compose one round. The entire procedure was repeated for 12 times (32 minutes) composing 12 rounds. In addition to the ten letters in alphabetical order, one round also contained the letters C, H and I at the end. Doing so, the number of letters stimulated in ABT is balanced with that of WBT where the letters C, H and I appear twice.

Reconstruction Testing (RT). Participants were asked to select (using the mouse or keyboard) which locations (vibromotors) are used to encode a given letter displayed on the screen. Figure 4 shows the user interface for RT.

Letter Testing (LT). Participants were stimulated with a pattern and asked to input the letter associated with it. They could repeat the stimuli before answering, and they were not notified whether their answer was correct.

Word Testing (WT). Like LT, users try to recognize stimulations. Participants were tested for words constructed only from letters that they trained. These include the four words used in WBT (ES, ICH, MAG, NICHT) and four additional words (IN, MIT, IST, GEHEN).

The first round of the sudy was a pre-test consisting of a round of RT and a round of LT. Pre-test served the purpose to familiarize participants with the testing procedure and to demonstrate their lack of knowledge about skin-reading. Thereafter, participants were exposed to the game and passive training. They were explicitly instructed to focus on the game. They were randomly assigned to two equal groups. The first group trained using WBT and the second one with ABT. After 32m of training, they were exposed to rounds of RT, LT and WT.

[1] `https://en.wikipedia.org/wiki/Snake_(video_game_genre)`

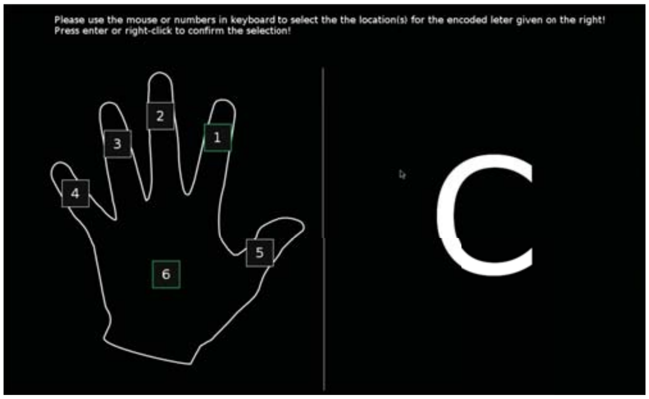


Figure 4: The user interface used for letter reconstruction test.

We refer to this block as post-train testing. To study the effect of transmission speed, post train LT and WT were performed with a base duration of $d \in [100ms, 200ms, 300ms]$ coupled, in the case of words, with between letter duration of $bl \in [100ms, 200ms, 300ms]$. Finally, participants filled out a NASA TLX questionnaire, rating workload of the letter and word recognition. They were also asked to rate the three following sentences using five-level Likert scale (from strongly disagree to strongly agree):
Effectiveness*: The (voice) passive training of letters while playing is a good way of teaching to recognise vibrotactile encoded letters!*
Annoyingness*: The (voice) passive training of letters while playing the game was annoying!*
Interruptedness*: The (voice) passive training of letters while playing the game did prevent me from focusing on the game!*

On the very next day, participants were exposed to another testing, identical to the post-train testing. We refer to it as recall-testing, as its purpose was to evaluate how much users recall the next day. The entire procedure is depicted in Figure 5. In the pre-train testing one probe per letter was collected in both LT and RT. During post-train and recall, for each letter, one probe was collected in RT, six probes (two per each speed) in LT and three probes (one per speed) WT.

Participants

Twenty (20) individuals (13 male and 7 female) aged between 23 and 46 (M=32.7, STD=7.6) years old participated in this study. Half of participants used WBT. Only one of them was left handed. All of them used the left hand for stimulation and the right to interact with the computer as depicted in Figure 3.

Results

Let us define the following variables: accuracy, repetition and total duration. Repetition describes how many times a user repeated the stimulation (letter or word) in LT, WT rounds. Total duration represents the difference between the user response time-point and the first stimulation time-point including repetitions.

Accuracy will be defined differently for different test types. For the RT accuracy is defined to be 1 if the user provides the exact locations that encode the given letter, otherwise 0. Similarly, for LT the accuracy is a binary variable defined to be 1 if the user's response matches the stimulated letter. For WT, accuracy is defined in relation to the similarity of the stimulated word to the user's response. Word recognition accuracy for a pair of answer and stimulated word (a,s) is computed by the given expression:

$$\sigma(a,s) = 1 - \frac{d(a,s)}{\#s}, \tag{1}$$

where d is the Levenshtein distance [8] between two words and $\#s$ represents the word length (number of letters). The Levenshtein distance is defined as the minimum single-letter edits (insertions, deletions or substitutions) required to change one word into the other [2].

We consider the testing phase (post-train, recall), speed (100 ms, 200 ms. 300 ms) and training method (ABT, WBT) as independent variables; the letter reconstruction accuracy, as well as recognition accuracy on word and letters as dependent variables. We also consider repetition rate and total duration dependent variables.

Letters

Table 2 lists letter reconstruction and recognition accuracies in the pre-train test. Participants managed to guess/identify letters with an accuracy of 6% and reconstruct them with an accuracy of 8% before the training. The results demonstrates their lack of knowledge about the encoding of letters. The letter recognition and reconstruction accuracies for the post-train and recall tests are presented in Table 1 and Table 3.

Considering that the recognition and reconstruction accuracy are binary values, we will use chi-squared analysis to determine the significance of differences in accuracy.

As regards RQ1, a chi-squared analysis revealed no significant difference in recognition accuracy between the post-train phase ($M = 0.69, STD = 0.46$) and recall ($M = 0.69, STD = 0.46$); $\chi^2(1, N = 2400) = 0.0, p = 0.96$. A chi-squared analysis reveals that there is no significant difference in reconstruction accuracy between the post-train phase ($M = 0.66, STD = 0.48$) and recall ($M = 0.66, STD = 0.47$); $\chi^2(1, N = 400) = 0.0, p = 1.0$ Furthermore, a chi-squared analysis reveals that the differences in accuracy between the recognition ($M = 0.69, STD = 0.46$) and the reconstruction of letters ($M = 0.66, STD = 0.48$) are not significant ; $\chi^2(1, N = 2800) = 1.21, p = 0.27$. We also explore the relationship between the letter recognition accuracy and the performance in the game while training which is presented in Figure 7. A Pearson correlation analysis reveals that there is no significant correlation between the average recognition accuracy and user's high-score in the game; $r = 0.35, p = 0.13$. Morover, Figure 7 clearly shows that the recognition accuracy varies a lot among users. There are 4 users that do not even achieve 40% accuracy. On the other hand, there is a cluster of 8 users that perform with an accuracy over 88% and the rest lie in between. Additionally, we explore the total duration from users' response. Since the values are neither binary nor normally distributed

[2] `https://en.wikipedia.org/wiki/Levenshtein_distance`

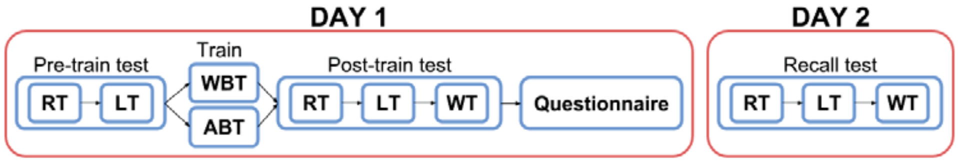
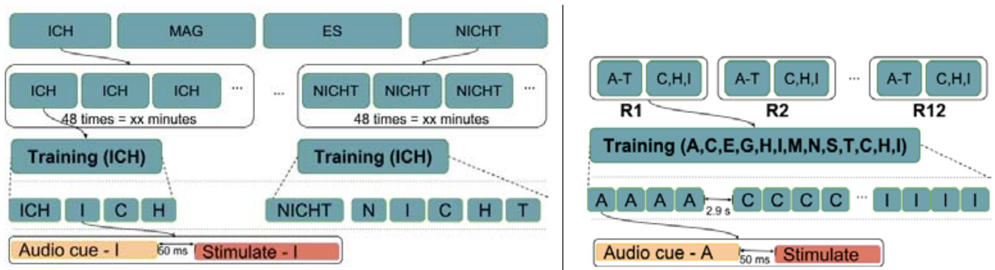


Figure 5: The entire procedure of PHL training and testing.



(a) Word based training (b) Alphabetical based letter training

Figure 6: Training methods: word based training (a) where letters from a word are used to determine the order of the trained letters and sequential training (b) where letters are trained in alphabetical order.

(Shapiro-Wilk test, $p < 0.05$), we rely on nonparametric tests for determining the significance. The effects of phase on duration until response are analyzed with Wilcoxon signed-rank test, as we are handling repeated measurements. The test reveals that indeed participants were **significantly** faster on the recall test ($MDN = 2.54, M = 3.8, STD = 3.88$) compared to the post-train ($MDN = 2.78, M = 4.17, STD = 4.29$); $V = 323582.5, p = 0.004$. Also repetition rate is non binary and not normally distributed and thus we will use the same nonparametric test. The test reveals that participant performed **significantly** less repetitions on post-train phase ($MDN = 1.0, M = 1.63, STD = 2.86$) than in recall ($MDN = 1.0, M = 1.95, STD = 3.23$); $r = 146372.0, p = 0.002$.

Regarding RQ2, there is no significant effect of speed in letter recognition accuracy; $\chi^2(2, N = 2400) = 0.0, p = 0.99$. We use Kruskal-Wallis to analyse the effect of transmission speed on duration until response. The test reveals that the duration is not significantly affected by the transmission speed; $H(2400) = 0.85, p = 0.654$. A Kruskal-Wallis test reveals taht the transmission speed had no effect on repetition rate; $H(2400) = 2.44, p = 0.295$.

As regards RQ3, there is a **significant** difference in recognition accuracy between participants trained with ABT ($M = 0.72, STD = 0.45$) and those trained with WBT ($M = 0.65, STD = 0.48$); $\chi^2(1, N = 2400) = 12.09, p < 0.001$. There is also a large, albeit non-significant difference in reconstruction accuracy between ABT ($M = 0.70, STD = 0.46$) and WBT ($M = 0.61, STD = 0.49$); $\chi^2(1, N = 400) = 3.6, p = 0.058$. To analyse how training method affects the response time we use Wilcoxon rank-sum test.The test reveals that the differences between WBT ($MDN = 2.72, M = 4.12, STD = 4.53$) and ABT ($MDN = 2.56, M = 3.84, STD = 3.6$) are not significant; $W = 1.39, p = 0.164$. Interestingly, participants trained using WBT did **significantly** fewer repetitions ($MDN = 0.0, M = 1.35, STD = 3.13$) than those trained with

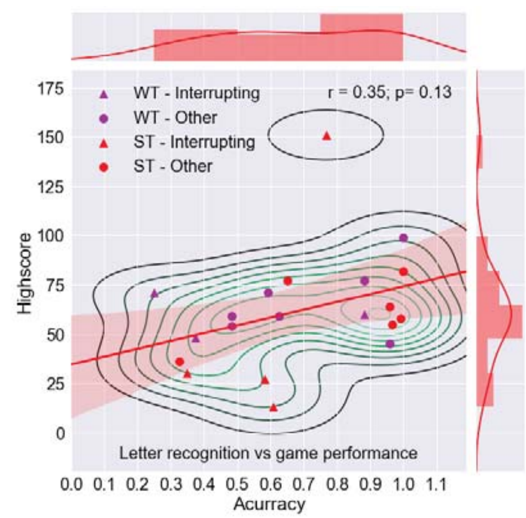


Figure 7: The relation between the average (per user) letter recognition accuracy and the user high score. The bar plots on the top and on the side represent histograms and calculated the univariate distribution of the variable in the given axis. The contours represent the multivariate distribution of both variables. The straight line and the shades around it represent the fitted regression and its confidence. The Pearson correlation index and the confidence value are annotated as r and p. The colour represents the training method whereas the shape expresses the participant's rating on the training experience.

ABT ($MDN = 1.0, M = 2.23, STD = 2.91$); $r = -12.46, p = 0.0$.

Words

The average word recognition accuracy, duration and repetition rate are presented in the Table 1. Given that the word

Phase	Train	Speed	Letters Accuracy	Letters Duration	Letters Repetition	Words Accuracy	Words Duration	Words Repetition
Post	ABT	100ms	.74 (.44)	3.94 (3.81)	2.30 (3.26)	.70 (.36)	6.52 (3.82)	6.35 (6.96)
		200ms	.71 (.45)	4.26 (4.56)	1.91 (2.88)	.69 (.34)	7.64 (5.89)	7.25 (10.3)
		300ms	.74 (.44)	4.05 (3.66)	1.64 (2.43)	.73 (.34)	6.17 (3.13)	4.94 (4.89)
	WBT	100ms	.68 (.47)	3.95 (3.59)	1.10 (1.74)	.64 (.41)	6.62 (3.94)	3.46 (3.01)
		200ms	.64 (.48)	4.59 (5.57)	1.61 (3.74)	.64 (.40)	6.40 (3.76)	3.65 (3.82)
		300ms	.62 (.49)	4.22 (4.21)	1.24 (2.57)	.71 (.36)	6.30 (3.32)	3.35 (3.78)
Recall	ABT	100ms	.70 (.46)	3.95 (3.56)	2.76 (2.98)	.72 (.33)	6.05 (3.22)	5.22 (5.45)
		200ms	.72 (.45)	3.30 (2.53)	2.44 (3.02)	.74 (.33)	5.61 (3.25)	5.29 (5.35)
		300ms	.72 (.45)	3.55 (3.09)	2.35 (2.72)	.75 (.32)	5.88 (3.78)	4.29 (4.49)
	WBT	100ms	.64 (.48)	3.74 (3.83)	1.46 (4.37)	.69 (.40)	5.93 (3.40)	3.76 (3.48)
		200ms	.68 (.47)	3.80 (4.15)	1.15 (2.29)	.69 (.39)	5.53 (2.89)	3.34 (3.01)
		300ms	.67 (.47)	4.44 (5.44)	1.52 (3.30)	.76 (.37)	5.51 (2.20)	2.56 (2.88)

Table 1: Letter and word recognition results (M,STD).

Method	LT	RT
ABT	.04 (.20)	.07 (.26)
WBT	.09 (.29)	.09 (.29)
Both	.06 (.25)	.08 (.27)

Table 2: Pre-train letter recognition and reconstruction accuracy (M,STD).

Phase	Train	Accuracy
Post	ABT	.70 (.46)
	WBT	.61 (.49)
Recall	ABT	.71 (.46)
	WBT	.61 (.49)

Table 3: Post-train and recall letter reconstruction accuracy (M,STD).

recognition accuracy is a real value and not normally distributed (Shapiro-Wilk test, p < 0.05), we rely on nonparametric tests (Kruskal-Wallis, Wilcoxon rank-sum and Wilcoxon signed-rank) for determining the significance.

As regards RQ1, a Wilcoxon signed-rank test reveals that indeed participants perform **significantly** better (accuracy) on the recall test ($MDN = 1.0, M = 0.72, STD = 0.36$) compared to the post-train ($MDN = 1.0, M = 0.68, STD = 0.37$); $V = 12990.5, p = 0.023$. Duration is not normally distributed (Shapiro-Wilk, p < 0.05). Comparing duration until response, a Wilcoxon signed-rank test reveals that participants were **significantly** faster on the recall test ($MDN = 5.05, M = 5.75, STD = 3.15$) compared to the post-train ($MDN = 5.39, M = 6.61, STD = 4.08$); $V = 45664.0, p = 0.0$. Participants also performed **significantly** more repetitions on post-train phase ($MDN = 3.0, M = 4.83, STD = 6.18$) than in recall ($MDN = 3.0, M = 4.08, STD = 4.33$); $V = 38467.0, p = 0.011$.

Regarding RQ2, A Kruskal-Wallis test reveals that the word recognition accuracy is not significantly affected by the transmission speed, $H(2) = 3.78, p = 0.15$. A Kruskal-Wallis test reveals that the duration is not significantly affected by the transmission speed; $H(960) = 1.36, p = 0.507$. A Kruskal-Wallis test reveals that the duration is not significantly affected by the transmission speed; $H(960) = 1.36, p = 0.507$. But, concerning repetition rate, the tests reveal that the transmission speed had a **significant** effect on repetition rate; $H(960) = 14.09, p = 0.001$. A further post-hoc Wilcoxon signed-rank tests reveal that participants did **significantly** fewer repetition when the vibrations and the gap between letters was set to 300 ms ($MDN = 2.0, M = 3.78, STD = 4.16$) compared to 200 ms ($MDN = 3.0, M = 4.88, STD = 6.48$); $V = 13594.5, p = 0.0$ and 100 ms ($MDN = 3.0, M = 4.7, STD = 5.1$); $V = 13944.0, p = 0.0$. However the differences between 200 ms and 100 ms were not significant; $V = 18676.0, p = 0.592$.

As regards RQ3, a Wilcoxon rank-sum test reveals that the differences in accuracy between WBT ($MDN = 1.0, M = 0.69, STD = 0.39$) and ABT ($MDN = 1.0, M = 0.72, STD = 0.34$) are not significant; $W = -0.63, p = 0.531$. Participants that were trained using WBT did **significantly** fewer repetitions ($MDN = 2.0, M = 3.35, STD = 3.36$) than the ones that used ABT ($MDN = 4.0, M = 5.56, STD = 6.6$); $W = -6.29, p = 0.0$.

Questionnaire

The users rating on the how effective the game based PHL is, how much it interrupts the game and whether it is annoying during the game, are visualised in Figure 8. The overall ratings are quite positive. However, there are a couple of participants that did provide some poor ratings. While the majority of the users thought that it is effective, two users disagreed, and four others were neutral. On the matter of being annoying, one user did find it annoying, and two others were neutral on this. Additionally seven users found it interrupting as they thought that the PHL did prevent them to focus on the game.

The results of NASA TLX for letter and word recognition tasks depending on the training method are depicted in Figure 9. We calculate the workload the simplified R-TLX method and compare the workload of letter and word recognition tasks between the training methods. Given that the workload values are normally distributed (Shapiro-Wilk: p > 0.05) and the variances of compared groups are homogenous (Levene: p > 0.05), we use the independent t-test. A t-test analysis reveals that the workload for letter recognition for participants that used ABT training method ($M = 4.22, STD = 1.47$) was lower that the workload of participants that trained used WBT ($M = 4.5, STD = 0.88$), but the differences are not significant; $t(20) = -0.52, p = 0.608$. When looking at the word recognition workload, on the contrary, participants that trained using ABT expressed a higher workload ($M = 5.68, STD = 1.4$) that participants that were trained using WBT ($M = 4.77, STD = 1.19$). However, again the differences are insignificant; $t(20) = 1.58, p = 0.13$.

DISCUSSION

Our user study was designed to investigate whether PHL could be used to train users for vibrotactile skin reading (RQ1) and explore different training methods (RQ3). Additionally, we investigate whether the transmission speed compared to the one that was used to train participants affects their ability to perceive the encoded information (RQ2).

The results of our study show that overall, both phases, both training methods and all speeds combined, participants

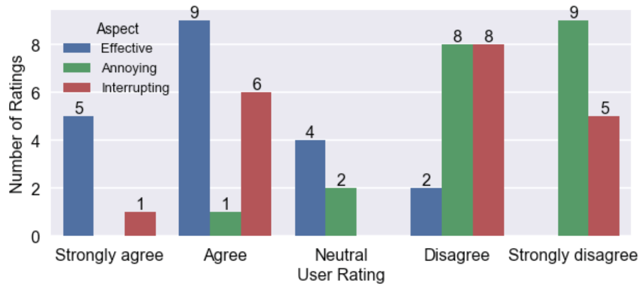


Figure 8: User ratings on how effective, interrupting and annoying the PHL is while playing the game.

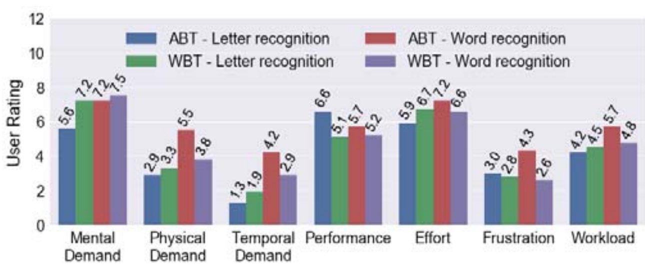


Figure 9: NASA TLX self evaluation metrics for letter and word and recognition tasks.

achieved a recognition accuracy of 69% on letters and 70% for words. To put this in perspective, Luzhnica and Veas [11] reported that 20 participants achieved an accuracy of 95% letter recognition after 5 iterations active training block each followed by a reinforcement block (in total 10 blocks). The reported accuracy is higher with less training blocks (12). Nevertheless, even though the learning rate is less than what participants could have potentially learned within the same time using active learning [11], recognising 69% of the letters using PHL means that in 32 minutes of training they were able to learn 7 letters in average. Thus in practice, for learning the entire Alphabet, one would need to reduce the number of letters within the 32 minutes of training and have more sessions (one per day) until users learn it entirely. Additionally, participants were able to not only recognise the letters but also reconstruct them which is consistent with the research on PHL of Morse code and Braille which showed successful reception along with reproduction [15, 19].

PHL has the benefit of letting user enjoying other activities while being trained and thus users would not need to stare at the screen and devote focus to training. Thus PHL could be used and presents an attractive alternative method for training vibrotactile skin-reading (RQ1). Whether, users would prefer spending more time in training but perform other activities during the same time (e.g. playing video games) or less time but focus actively on training, or even mix them, it would be up to individual preferences and should be considered as a trade-off. Moreover, six users did achieve an accuracy of or close to 100% (see Figure 7), meaning that for them, 32 minutes of PHL training was enough to learn 10 letters. Others demonstrated less learning from PHL. While with this study we were unable to explain their poor performance, we will explore such phenomena in the future and investigate whether by tuning training parameters such as the time from the sound cue to vibrotactile stimulation, the volume, personalised training method (e.g. different number of letters for different users within the same time) etc... to improve the learning effect for such users.

We also investigate whether there is a trend that users who did well at learning also did poorly at the game, suggesting that they possibly attended to the stimuli actively. However, we found no such trend and those who did better at learning were also some of the ones who did best at the game.

Considering the results of the word based (WBT) and alphabetically based training (ABT) methods (RQ3) demonstrated that both methods could be used for training. Nevertheless, findings on learning condition differences were surprising. Results suggest that the ABT condition enabled significantly better recognition and production of letters, and comparable performance on words; though this group required (significantly) more repetitions. This would indicate that perhaps the ABT allowed comparable learning while also helping users think of letters individually rather than strongly tied to their word. The research team expected that learning from the disorganised ABT condition would be very challenging and that the cognitive benefit from semantic associations and small groups of letters in the WBT condition would allow those users to perform significantly better. Given the surprising results which suggest the promise in the ABT condition, this work clearly shows that further consideration of the ABT vs WBT learning structure is needed.

The analysis of the recognition in different phases (post-train and recall) show that participants were able to recall the learned information after one day. Retention and recall is well known for traditional learning methods; however, it is often asked but still unknown whether learning from PHL lasts. The breadth of passive tactile learning research - from piano to rehab - has yet to explore this important question[15, 17, 19]. Further research should investigate later recall tests in different PHL scenarios, but this initial result is encouraging that the effects of PHL are beyond short-term working memory. Moreover, our results show an improvement on the subsequent day. Perhaps performance improved after a night of sleep or a break, which litterature suggests aids motor learning [22, 26].

Last but not least, our results show that participants were able to comprehend the transmitted information with the same accuracy regardless of the transmission speed (RQ3). This is consistent with related work which showed stimuli of different durations could be equally recognised on the fingers actively [18], assuming that the minimum duration threshold has been considered. Moreover, the results also suggest that the duration of stimuli could be decoupled between training and the use of the device, meaning that participants could train with one speed and once they learn the Alphabet and the use the device with faster speeds or even adjust the speed during usage without re-training.

CONCLUSION

This paper investigates the potential of passive haptic learning (PHL) as a training tool for vibrotactile skin reading. A user study puts 20 participants through a 32 minutes PHL training while they are actively engaged in playing a video game. The testing of the recognition of letters and words shows when trained, participants could recognise letters with an average accuracy of 69% and words with an accuracy 70%. Additionally, our study shows that PHL can be used regardless of whether the training is based on semantically grouped letters or alphabetically ordered ones. Moreover, the results show that participants recognitions accuracy was not affected by transmission speed indicating that they could be trained with a default speed and then proceed to use the system in different other levels of speed without requiring a re-training. Overall our results demonstrate that PHL presents an alternative to active learning for training vibrotactile skin reading.

ACKNOWLEDGMENTS

This material is based upon work supported, in part, by MoreGrasp (H2020-ICT-2014-1 643955) and by a Microsoft Research PhD Fellowship and the National Science Foundation under grant No. 1217473. The Know-Center is funded within the Austrian COMET Program - Competence Centers for Excellent Technologies - under the auspices of the Austrian Federal Ministry of Transport, Innovation and Technology, the Austrian Federal Ministry of Economy, Family and Youth and by the State of Styria. COMET is managed by the Austrian Research Promotion Agency FFG. This material is based upon work supported, in part, by a Microsoft Research PhD Fellowship and the National Science Foundation under grant No. 1217473.

REFERENCES

1. J. C. Bliss, M. H. Katcher, C. H. Rogers, and R. P. Shepard. 1970. Optical-to-Tactile Image Conversion for the Blind. *IEEE Transactions on Man-Machine Systems* 11, 1 (March 1970), 58–65. DOI: **http://dx.doi.org/10.1109/TMMS.1970.299963**
2. Robert O. Duncan and Geoffrey M. Boynton. 2007. Tactile Hyperacuity Thresholds Correlate with Finger Maps in Primary Somatosensory Cortex (S1). *Cerebral Cortex* 17, 12 (2007), 2878. DOI: **http://dx.doi.org/10.1093/cercor/bhm015**
3. Robert H Gault. 1924. Progress in experiments on tactual interpretation of oral speech. *The Journal of Abnormal Psychology and Social Psychology* 19, 2 (1924), 155.
4. Frank A. Geldard. 1957. Adventures in tactile literacy. *American Psychologist* 12, 3 (1957), 115–124.
5. Eric Gunther. 2001. *Skinscape: A tool for composition in the tactile modality*. Ph.D. Dissertation. Massachusetts Institute of Technology.
6. Eve E. Hoggan, Sohail Anwar, and Stephen A. Brewster. 2007. *Mobile Multi-actuator Tactile Displays*. 22–33 pages. DOI: **http://dx.doi.org/10.1007/978-3-540-76702-2_4**
7. Jacob H. Kirman. 1974. Tactile perception of computer-derived formant patterns from voiced speech. *The Journal of the Acoustical Society of America* 55, 1 (1974).
8. VI Levenshtein. 1966. Binary Codes Capable of Correcting Deletions, Insertions and Reversals. *Soviet Physics Doklady* 10 (1966), 707.
9. Yi-Chi Liao, Yi-Ling Chen, Jo-Yu Lo, Rong-Hao Liang, Liwei Chan, and Bing-Yu Chen. 2016. EdgeVib: Effective Alphanumeric Character Output Using a Wrist-Worn Tactile Display. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. DOI:**http://dx.doi.org/10.1145/2984511.2984522**
10. Granit Luzhnica and Eduardo Veas. 2017. Vibrotactile Patterns using Sensitivity Prioritisation. In *Proceedings of the 2017 ACM International Symposium on Wearable Computers (ISWC '17)*. ACM, New York, NY, USA. DOI:**http://dx.doi.org/10.1145/3123021.3123032**
11. Granit Luzhnica and Eduardo Veas. 2018. Investigating Interactions for Text Recognition Using a Vibrotactile Wearable Display. In *23rd International Conference on Intelligent User Interfaces (IUI '18)*. 453–465. DOI:**http://dx.doi.org/10.1145/3172944.3172951**
12. Granit Luzhnica, Eduardo Veas, and Viktoria Pammer. 2016. Skin Reading: Encoding Text in a 6-channel Haptic Display. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers (ISWC '16)*. ACM, New York, NY, USA. DOI: **http://dx.doi.org/10.1145/2971763.2971769**
13. Hugo Nicolau, João Guerreiro, Tiago Guerreiro, and Luís Carriço. 2013. UbiBraille: Designing and Evaluating a Vibrotactile Braille-reading Device. In *International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '13)*. DOI: **http://dx.doi.org/10.1145/2513383.2513437**
14. Scott D. Novich and David M. Eagleman. 2015. Using space and time to encode vibrotactile information: toward an estimate of the skin's achievable throughput. *Experimental Brain Research* 10 (2015). DOI: **http://dx.doi.org/10.1007/s00221-015-4346-1**
15. Caitlyn Seim, John Chandler, Kayla DesPortes, Siddharth Dhingra, Miru Park, and Thad Starner. 2014. Passive Haptic Learning of Braille Typing. In *ACM International Symposium on Wearable Computers (ISWC '14)*. DOI:**http://dx.doi.org/10.1145/2634317.2634330**
16. Caitlyn Seim, Nick Doering, Yang Zhang, Wolfgang Stuerzlinger, and Thad Starner. 2017. Passive Haptic Training to Improve Speed and Performance on a Keypad. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 100 (Sept. 2017), 13 pages. DOI: **http://dx.doi.org/10.1145/3132026**
17. C. Seim, T. Estes, and T. Starner. 2015a. Towards Passive Haptic Learning of piano songs. In *2015 IEEE World Haptics Conference (WHC)*. DOI:**http://dx.doi.org/10.1109/WHC.2015.7177752**
18. Caitlyn Seim, James Hallam, Shashank Raghu, Tri-An Le, Greg Bishop, and Thad Starner. 2015b. *Perception in Hand-Worn Haptics: Placement, Simultaneous Stimuli, and Vibration Motor Comparisons*. Technical Report. Georgia Institute of Technology.
19. 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)*. ACM, New York, NY, USA. DOI:**http://dx.doi.org/10.1145/2971763.2971768**
20. 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*. ACM. DOI:**http://dx.doi.org/10.1145/2559206.2581329**
21. Annette Sterr, Lisa Green, and Thomas Elbert. 2003. Blind Braille readers mislocate tactile stimuli. *Biological Psychology* 63, 2 (2003). DOI:**http://dx.doi.org/10.1016/S0301-0511(03)00051-6**
22. Robert Stickgold and Matthew P Walker. 2013. Sleep-dependent memory triage: evolving generalization through selective processing. *Nature neuroscience* 16, 2 (2013), 139.
23. Ian R Summers, Jon J Whybrow, Denise A Gratton, Peter Milnes, Brian H Brown, and John C Stevens. 2005. Tactile information transfer: A comparison of two stimulation sites. *The Journal of the Acoustical Society of America* 118, 4 (2005), 2527–2534.
24. David Ternes and K E MacLean. 2008. Designing Large Sets of Haptic Icons with Rhythm. In *EuroHaptics*, M. Ferre (Ed.). Springer.
25. Francisco Vega-Bermudez and Kenneth O Johnson. 2001. Differences in spatial acuity between digits. *Neurology* 56, 10 (2001), 1389–1391.
26. Matthew P. Walker, Tiffany Brakefield, Alexandra Morgan, J.Allan Hobson, and Robert Stickgold. 2002. Practice with Sleep Makes Perfect: Sleep-Dependent Motor Skill Learning. *Neuron* 35, 1 (2002), 205 – 211. DOI:**http://dx.doi.org/10.1016/S0896-6273(02)00746-8**
27. Benjamin W. White, Frank A. Saunders, Lawrence Scadden, Paul Bach-Y-Rita, and Carter C. Collins. 1970. Seeing with the skin. *Perception & Psychophysics* 7, 1 (1970), 23–27. DOI: **http://dx.doi.org/10.3758/BF03210126**

28. Cheng Xu, Ali Israr, Ivan Poupyrev, Olivier Bau, and Chris Harrison. 2011. Tactile Display for the Visually Impaired Using TeslaTouch. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*. DOI:http://dx.doi.org/10.1145/1979742.1979705

29. Siyan Zhao, Ali Israr, Frances Lau, and Freddy Abnousi. 2018. Coding Tactile Symbols for Phonemic Communication. In *CHI Conference on Human Factors in Computing Systems*. DOI: http://dx.doi.org/10.1145/3173574.3173966