Learn with Haptics: Improving Vocabulary Recall with Free-form Digital Annotation on Touchscreen Mobiles

Smitha Sheshadri[1], Shengdong Zhao[1], Yang Chen[1,2], Morten Fjeld[3]
[1]NUS-HCI Lab, National University of Singapore, Singapore
[2]Zhejiang University, Hangzhou, China
[3]t2i Lab, Chalmers, Gothenburg (S) and University of Bergen (N)
smitha3shesh@gmail.com, zhaosd@comp.nus.edu.sg, sonnechen@zju.edu.cn, fjeld@chalmers.se

ABSTRACT
Mobile vocabulary learning interfaces typically present material only in auditory and visual channels, underutilizing the haptic modality. We explored haptic-integrated learning by adding free-form digital annotation to mobile vocabulary learning interfaces. Through a series of pilot studies, we identified three design factors: annotation mode, presentation sequence, and vibrotactile feedback, that influence recall in haptic-integrated vocabulary interfaces. These factors were then evaluated in a within-subject comparative study using a digital flashcard interface as baseline. Results using a 84-item vocabulary showed that the 'whole word' annotation mode is highly effective, yielding a 24.21% increase in immediate recall scores and a 30.36% increase in the 7-day delayed scores. Effects of presentation sequence and vibrotactile feedback were more transient; they affected the results of immediate tests, but not the delayed tests. We discuss the implications of these factors for designing future mobile learning applications.

Author Keywords
Motoric engagement; Mobile vocabulary learning; Haptics for learning; Multimodal learning; Intersensory reinforced learning.

CCS Concepts
•**Human-centered computing** → **Touch screens; Empirical studies in HCI;** *HCI theory, concepts and models;*

INTRODUCTION
Vocabulary acquisition is fundamental to learning a second language. Fluency in more than one language is regarded as a highly marketable skill [19]. In the diverse and connected world we live in today, languages are aptly called *passports to the world* as they enable inter-cultural immersion and collaboration. Vocabulary learning gets a significant boost from mobile devices. These devices allow learners to perform self-learning anytime, anywhere, which can significantly enhance their exposure to the learning content [23, 15]. Additional benefits of mobile vocabulary learning include the adoption of bite-sized lessons [48] so that the learning content can become more manageable and enriched with multi-modal elements to improve recall [64].

However, current mobile vocabulary learning applications typically only leverage visual and auditory modalities, neglecting the potential benefit of haptic elements in learning. Here, we use the term *haptic elements* to refer to the following two components: 1) users' finger movements on the screens or the passive haptic element [37]; 2) vibrotactile feedback (VTF) from a mobile device [45]. Haptic elements can be employed to achieve motoric engagement, as favored by cognitive theory to be the basis of cognition and understanding [7, 25, 72].

Here, we explore the possibility of improving recall of vocabulary items through motoric-engagement with digital annotations that are created with finger movements on touchscreens. We design the setting with an iterative approach to tease out various design factors that influence recall in a motorically-reinforced vocabulary presentation setting for mobile screens.

We addressed the decision on the stimulus or the prompts learners should annotate by experimenting with several candidates in a pilot study. We also investigated the integration of the haptic element into the existing presentation settings so that the various modal elements reinforce each other to help recall. We explored two alternative temporal relationships between the different modal elements. Additionally, we examined the efficacy of vibrotactile feedback (VTF), which is a widely used haptic capability in mobile phones.

We propose these three factors of annotation stimulus, presentation sequence and VTF as design factors which affect recall of vocabulary items. To validate these factors, we designed a within-participant study with an 84-item vocabulary and 12 participants, conducted sessions for 3 days and recorded the performances in tests of vocabulary recall.

Our results show that including annotation significantly improves recall of vocabulary items, with a mean increase of 30.36% compared to the control condition with 28 vocabulary items in each condition in cued tests. We also discovered that certain combinations of the design factors are more suitable to specific application cases as the results showed a significant interaction between the annotation mode and presentation sequence. Our results also show that VTF can significantly improve the free recall of vocabulary in immediate tests.

CHI '20, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6708-0/20/04 ...$15.00.
http://dx.doi.org/10.1145/3313831.3376272

This paper has two main contributions: We identify three design considerations for a motorically-reinforced vocabulary learning setting by analysing the problem from an interdisciplinary perspective of cognitive theories and Human Computer Interaction. Secondly, we validate the efficacy of these factors with an empirical study showing significant improvement in vocabulary learning while using existing hardware/software. Further, while our focus is vocabulary learning, the implications potentially have wider applicability to other paired-associative learning tasks.

BACKGROUND AND RELATED WORK

Improvements to vocabulary learning

In this section, we group the approaches employed to improve vocabulary learning in mobile phones into three categories. First, several studies have leveraged the anytime-anywhere capability of mobile phones. Studies have proposed context-dependent systems which present new vocabulary items relevant to the learner's physical environment [74, 32, 8, 20]. Context-dependent presentation has been reported to have a greater effect on gaining knowledge of meaning than increasing number of exposures [70]. Other studies have attempted to utilize wait-times in a learner's day and present vocabulary [10, 20].

Second, studies have proposed models which adapt to individual learners. Kim et al. [31] estimate a learner's level of proficiency, uses their Facebook feed to analyse their interests and customize lessons accordingly. Chen and Chung [14] present vocabulary customized to learner's progress and memory cycles. The spacing between word exposures, testing intervals and other factors of vocabulary presentation can be adapted based on individual learner's progress [58, 20].

The third category involves studies that have attempted to improve learning by enriching the presentation framework with multi-modal content. Pemberton et al. [46] facilitate the collection and tagging multimedia content like text and images learners come across in their everyday life to help them progress in language learning. Lin and Yu [36] experiment with different presentation modes involving text, audio and picture and report significant improvements by adding audio to vocabulary items. Pires et al. [47] analysed smartphone applications for vocabulary learning and regard images as an integral element of the applications. Pu and Zhong [50] propose a mobile augmented reality game for preschool students to present rich material.

Although the third category of studies propose to enrich the presentation material with multi-modal content, to our knowledge, no study has examined the integration of motoric modality to improve recall of vocabulary on mobile phones, despite there being substantial evidence in material medium that motoric encoding improves learning.

Motoric engagement and learning

Motoric images are elemental in language comprehension as it promotes *the storage of some type of motoric trace or image* [57]. Relevant engagement of the motoric system through locomotion, kinaesthetic motions or fine movements create motoric images. We can typically see studies take one of two approaches to achieve motoric-engagement; either through bodily actions like enactment [56], gestures [38, 39, 52] and manipulation of physical objects [24, 41, 54], or through fine motor activities like finger tracing, finger gestures, drawing and writing. A validated taxonomy of motoric-engagement [61] suggests four degrees of motoric engagement (the fourth one being the most immersive). They classify any activity which involves small movements on a touchscreen or mouse-driven movements on a desktop monitor or tablet with generative interactivity as second degree motoric engagement.

In our context, we will refer to this as simply motoric engagement. We use the term motorically reinforced platform to refer to our proposed platform for vocabulary presentation which facilitates motoric engagement. Next, we elaborate on approaches using second degree motoric engagement.

Second degree motoric engagement activities

Wammes et al. [68, 69] report that drawing improves the recall of words and also suggest that drawing is better for recall than writing. We will address this comparison in detail in the next sections. Agostinho et al. [1] investigate the benefits of using *biologically primary knowledge* of finger tracing on temperature graphs to help learners understand the graphs better and report better performance for the tracing conditions. Dubé and McEwen [18] employ drag gestures and report better performance in number-line estimation tasks using gestures. It reasons that drag gestures have a sense of continuity in them which reinforced the continuous understanding of numbers.

Literature stands divided on the benefits of fine motor activation with writing. Several studies have reported an improvement in recall of words which are written [44]. However, Barcroft [6] argues that writing is a form of production without access to meaning and it actually impedes learning, reporting a study that made learners attempt to acquire new vocabulary by either writing the whole word, writing a word fragment or without writing. The study reported that the writing conditions performed sub-optimally compared to the no writing conditions. Barcroft [6] uses a hypothesis proposed in a previous work [5] to explain the sub-optimal performance for write conditions. On the other hand, the writing superiority effect, a well-received theory in cognition propagates that writing allows for better information encoding and thus helps learning and recall [26].

In our study, we explore motoric-engagement with fine motor activities through writing. We attempt to leverage the haptic capabilities of touchscreen devices and engage the learner in relevant motoric-engagement through writing while learning vocabulary. We also analyse how the haptic modality would combine with the other modal elements in a vocabulary presentation framework.

RESEARCH OBJECTIVE

Throughout this paper, free-form digital annotation refers to any note, verbal (written) or diagrammatic, a user makes on their digital device without geometric constraint [62]. The primary focus of this study was to investigate whether the incorporation of free-form digital annotation into a mobile

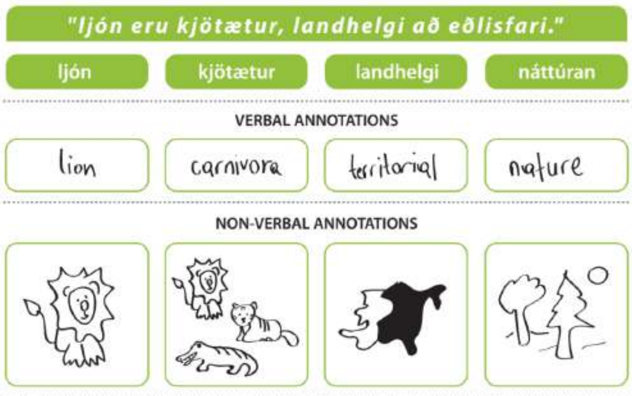

Figure 1. Examples of verbal (top) and non-verbal (bottom) annotations a learner produces for a sentence in Icelandic (Íslenska).

learning platform improves recall of foreign vocabulary. The primary decision was regarding what to annotate (the annotation stimulus).

Annotation stimuli

Annotation stimulus is the prompt learners should reproduce grapho-motorically. Annotation stimuli can be broadly classified into two categories, non-verbal (diagrammatic) and verbal (written) [2]. In our study design, learners achieve motoric engagement by reproducing the annotation stimuli on a touch-screen device with the index finger of their dominant hands.

Non-verbal or diagrammatic stimuli are image annotations corresponding to an object or concept. Levin [35] describes several functional benefits of image illustrations or verbal stimuli in information comprehension. Verbal stimuli are words written in a language to describe or represent an object or a concept. Figure 1 illustrates the comparison between the two stimulus classes. For our design, we chose to address a set of three requirements.

Relevance: The stimulus had to be relevant to the learning material. Study designs have worked with motoric engagement that is not directly relevant to the learning material and have failed to extract any benefits [65].

Scalability across a complete vocabulary: Words in any language vary widely in length, part of speech, specificity, multiplicity of meaning [33] and any comprehensive design should cater to learning these words. Wammes et al. [67] recommend that drawing is more beneficial than writing. However, drawing take up considerably more time than writing [67]. Also, finding visual prompts for all words is an immensely difficult task. Figure 1 shows a comparison between the verbal and visual stimuli. It is clear that visual stimuli become increasingly difficult to reproduce and ambiguous (lion->territorial).

Scalability across all learners and elimination of training overhead: Visual illustration or drawing skills differ from person to person. We aimed at creating a setting which would cater to all learners without any need for training.

These requirements lead us to use verbal prompts. A vocabulary item of the form "L2 word : L1 meaning (i.e., ljon : lion)" has two elements. The second language word or the L2 word (ljon) is the new word a learner is being introduced to and the L1 word (lion) informs the learner of the meaning through a language known to them. The next question was, do learners need to write both L1 and L2 elements to benefit their recall? Would annotation of the newly introduced L2 word suffice? To address such questions, we designed pilot study 1; we used flashcards as the baseline for comparison as these are the most commonly used method by students to memorise information [73]. We describe the vocabulary items and our choice for first language and second language in the later sections.

Abbreviation	Full form
Ww	Write whole word
Ws	Write word skeleton
nW	No write
Seq	Sequential presentation
Con	Concurrent presentation
VTF	Vibrotactile feedback
no VTF	No Vibrotactile feedback
L1	First language
L2	Second language

Table 1. List of abbreviations used. Pilot and participant labels are excluded.

Pilot study 1 (p1)

Objective: Compare the test performance yielded by training using three levels of annotations: writing of both L1 and L2 words (*both*), writing of L2 word only (*L2*) and flashcards (*control or no write*)

Scope: Five participants (P1-P5) were trained in a 90 item vocabulary for 3 sessions over 3 days. None of the participants were repeated in any two pilots or in a pilot and the final study. Each session had three subsets for each condition. Participants were given instructions about what to write for each subset. After the session on day three, participants were given the assessment with all L1 words and were asked to recall and write the L2 word. There was no time limit given.

Application: The application was designed on Android platform. There was an audio pronunciation of the L2 word generated using text to speech software and could be listened to by clicking the audio button. Participants could click on the next button to proceed to the next word.

Results: (Maximum score of 30 per condition)

The mean scores for the three conditions were 18.0, 15.2 and 10.2 as shown in figure 2 for *both*, *L2* and *control* respectively. Results indicate that there was an improvement to the performance in recall tests when either both elements of the L1-L2 pair or only the L2 word is written. Although condition *both* performed better than presenting *L2* only, interviews with the participants revealed that writing only the L2 word was preferred over writing the L1 meaning which corresponded to very well known objects.

Participant feedback and analysis: Participants prefers using only the annotations of the L2 word. Three of the five participants reported difficulties using their mobile screen to

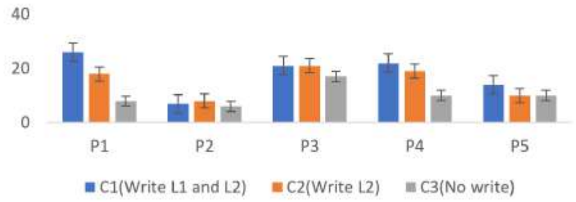

Figure 2. Pilot 1: Mean scores obtained by participants out of 30.0 per condition C1-C3.

write longer words. P4 remarked, *"Writing entire word becomes a little tedious. And because it [the phone screen] is not very big, I also have to worry about making sure there is enough space for entire word."* We decided to investigate this problem further and our considerations were: Using the landscape orientation in phones to accommodate long words might appear to be a straightforward solution. However, landscape mode is less preferred by mobile users [60]. Furthermore, this solution would not address the scalability of the design. For instance, words could get too long even for landscape orientation.

We also analysed the first part of the users' comment, about how writing long words can be tedious. While this is a subjective opinion expressed in a limited scope interview, we could upgrade the annotation design by exploring alternatives for whole word annotation using less space on screen and demanding less work from the learner.

Driven by this motivation, we decided to investigate the effects of shortening the L2 word or using word skeleton annotation and compare it with whole word annotation.

Word skeleton annotation

In order to find a shorter alternative for the L2 annotation, we analysed the work on word shortening [53] and the role of vowels and consonants [12] in recognising English words. It has been shown that participants have a harder time recognising words missing consonants than those with missing vowels [22]. Furthermore, according to Rawlinson [51] inaccuracies in the order of letters do not significantly affect readers as long as the first and last letters are unchanged. We borrowed from these ideas and used the following algorithm to create word skeletons from the L2 words. Analysis of screen recordings from the participant sessions of pilot 1 showed that words with less than 5 characters would fit easily into the screen, we thus employed word shortening only for words with 5 characters or more. The algorithm proposed was: If the word was shorter than 4 characters, we used it as it is. If the word was at least 5 characters long, the first and last characters were unchanged.The vowels were dropped, if there were consecutive vowels, only first vowel was dropped. For example, Malbex : Mlbx, Cosour : Csur, Viggel : Vggl.

Implications: i1. Writing of the L2 word might benefit recall of vocabulary items more than no writing. i2. Writing of the shortened form of the L2 word might perform as well as writing the whole L2 word. Next, we address the role of the new haptic modality concerning the other modal elements.

Temporal relationship between the modal elements

According to the multimedia theory of learning by Mayer [42], humans process different modal inputs separately and connections between these representations are necessary for meaningful learning. The modal elements involved in our design were; the image element, the verbal elements (the L1 word and the L2 word), the auditory element (L2 pronunciation) and the haptic element. The haptic element can be broken down into two separate modalities, the motoric input and the verbal output. The motoric input is when the learner expends effort in creating the strokes. These strokes in turn form letters which make up the word reproducing the verbal prompt. Multimedia theory also states that each modal channel has limited bandwidth which needs to be utilized carefully while presenting information. In our design, we could either present all modal elements concurrently or introduce a temporal sequence in which each modal element appears in succession. The first approach is likely to encourage the active-processing assumption of multimedia learning; the second might be more resistant to potential cognitive overload. We designed pilot study 2 with a goal of comparing these approaches.

Pilot study 2 (p2)

Objective: Compare the test performance yielded by two conditions of presentation layout: sequential presentation based on haptic input (*Seq*) and concurrent presentation (*Con*) along with whole L2 word annotations.

Scope: Four participants (P1-P4) were trained in a 40 item vocabulary for 2 sessions spanned across 2 days.

Design: The 40 vocabulary items were divided into 2 sets for *Seq* and *Con*. Words corresponding to each condition were ordered randomly and participants had 2 sessions of 40 items. The test of all 40 items were given after session 2, following a 5 minute break. In the test, a list of all L1 words were given and participants had to recall and write the L2 word. There was no time limit given.

Layout: For *Seq* condition, the layout was initially empty except for the writing stimulus (L2 word). Once the participant starts to write, the strokes were recorded and the image along with the L1 word were faded in by gradually increasing the opacity. A diagrammatic representation of this flow can be seen in figure 4. In *Con* condition, the layout appeared with all the elements as seen in figure and remained static but for the writing. Sequential presentation based on haptic input (Seq) outperformed concurrent presentation (Con). Participants remarked that sequential presentation allows them to fully concentrate on one element and then move onto the next. P1 mentioned, *"I like the fading in, it allows me to take some time to read that word* [L2 word] *and then deliberately get the meaning."* These remarks and the promising trends shown by this pilot as shown in figure 3 encouraged us to retain presentation sequence as a design factor.

Implications: i3. Sequential presentation based on haptic input (*Seq*) of the different modal elements might help recall better than concurrent presentation (*Con*).

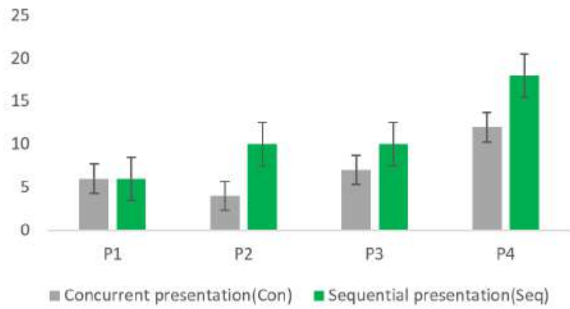

Figure 3. Pilot 2: Mean scores obtained by participants out of 20.0 per condition P1-P4.

Haptic experience and Vibrotactile feedback (VTF)

Brewster and Brown [9] advocate for the incorporation of cutaneous perception into any haptic interface to enhance its usability and effectiveness. They reason that touch has two integral elements- *the kinaesthetic perception* which processes the *information arising from forces and positions sensed by the muscles* and *the cutaneous perception* which caters to sensations such as vibration, temperature etc. Also, VTF has been shown to increase the feel of writing, participants responded positively to vibrotactile cues from the writing surface and reported that it felt comfortable and resembled the feel of writing in a material medium [16]. Therefore, we designed the next pilot study to investigate how VTF is received by learners as a tool to increase familiarity of the setting.

Pilot study 3 (p3)

Objective: To explore whether VTF during annotation improves user experience by making the setting more similar to the familiar experience of writing on a material medium.

Scope: 5 participants (P1-P5), exposure to 10 vocabulary items, 5 were chosen randomly for each participant for set 1 (VTF) and set 2 (no VTF)

Layout: Words with VTF vibrated on touch. Participants were interviewed after the session.

Interview questions: Did you notice any difference between words of set 1 and set 2? Which would you prefer to digitally annotate? Why?

Participant feedback: Participants were mostly positive towards VTF (3 out of 5 preferred writing with vibration and 1 had no preference), P4 remarked, *"Vibration just feels better, once I'm used to writing with vibration, the one without feels weird."* P1 mentioned that *"It feels more like actual writing, but I'm not really against the one without either."*

Implications: i4. VTF might help improve user experience by increasing familiarity with a writing setting.

Based on the results of the three pilots, we wrapped up with the following three design factors for further investigation.

1) Annotation mode, 3 levels: Write whole word (*Ww*), Write word skeleton (*Ws*), No write (*nW*). The write whole word level involves the writing of the complete L2 word. The word skeleton annotation is a potential alternative we investigate. If the results of the word skeleton annotation are comparable to whole word annotation, it could be used to save time and effort and enable scaling of the design to cover long words. Also, it could indicate that digital annotation can cater to concepts that have to be shortened to be easily annotated. The No write level is designed to function as the baseline representing flashcards.

2) Presentation sequence mode, 2 levels: Presentation in sequence based on haptic input (*Seq*), Concurrent presentation (*Con*). Presentation in sequence based on haptic input initially presents only the writing stimulus in an otherwise empty layout. The participant then has to start writing the stimulus. The strokes are captured and the image element and the L1 verbal element are slowly revealed to the participant by introducing a direct proportionality between the completeness of the written word and the visibility of the image and L1 verbal element. On the other hand, the concurrent presentation has all the elements present already and the participant needs to write the stimulus on an otherwise static layout. The concurrent presentation level serves as the baseline when combined with the No write level being identical to the flashcard design.

3) Haptic feedback mode, 2 levels: vibrotactile feedback (*VTF*) and no vibrotactile feedback (*no VTF*): VTF gives feedback in the form of vibration as the participants write. The no VTF condition is the baseline as it represents the flashcard design. We point out here that this design does not explore some dimensions of VTF such as frequency and duration of the vibration as done by studies conducted by Brewster and Brown [9], however, our design isolates the effects of the presence haptic feedback on recall performances.

In addition to shedding light on the above design considerations, our pilots uncovered some critical observations we needed to address: Participants form word associations to help them link the L1 and L2 words. This is a widely used strategy in new vocabulary acquisition known as the keyword method [55].

Some participants expressed that they had more time to form such associations for the control group. We needed to address this issue as it introduced some variability between the conditions. Further analysis and discussions with participants revealed that these associations are typically formed upon the first exposure to the L1-L2 pair. Subsequent exposures do not call for new associations. Implementing the annotation conditions for the first exposure might affect the formation of these associations variably across the different conditions.

To avoid this confounding factor, for the final study, we decided to have uniformity in first exposure by having an initial common session in which all words were presented with the flashcard layout. The conditions were implemented from the subsequent sessions i.e. the revision sessions.

In pilot 1 and 2, participants controlled when to proceed to the next word by clicking on the next button. We recorded the time a participant spends on each word- 8.25s for the words with L2 writing and 7.85s for the words with no writing. To ensure uniformity in exposure time for all words, the next button was removed and the time for each word was set to be

8s for each No write word. For the write whole word and write word skeleton words, progress depended on the writing. We also recorded the number of times participants use the audio pronunciation. For most words, they used the pronunciation twice. We thus assigned two pronunciations of the L2 word per every vocabulary item in a session.

RESEARCH QUESTIONS AND HYPOTHESES

Based on the implications of the pilot studies, we formulate four research questions and five hypotheses:

RQ1: How does the adding whole and partial word annotations to the revision phase of vocabulary learning affect recall?

H1.1: Vocabulary items learnt by revising with annotations will be recalled better in the immediate and delayed cued and free recall tests than those revising using digital flashcards.

H1.2: There will be no significant differences between the performance of partial and whole word annotations in immediate or delayed cued or free recall tests. We hypothesize that annotating with the L2 word will lead to significant improvement in recall of vocabulary items. Additionally, we also attempt to upgrade this design factor by using partial word annotations formed using the algorithms based on linguistic evidence of word shortening. We hypothesize that shortening the words will not cause any significant impediment to the benefits of word annotations.

RQ2: What is the relationship between temporal sequence of presentation of various modal elements and recall of vocabulary items?

H2: Presentation in sequence will perform better than concurrent presentation in immediate and delayed cued and free recall tests. Competition between elements for modal bandwidth has been shown to deter learning. For example, there is a significant decrease in performance when scientific explanation are presented as animation and on-screen text (both target visual modality) than as animation (visual) and narration (auditory) [43]. We hypothesize that integrating the haptic modality to a sequential presentation environment will be more favourable to recall than a presentation environment without any sequence of presentation.

RQ3: How does adding vibrotactile feedback to writing affect recall of vocabulary?

H3: VTF improves immediate and delayed cued and free recall of vocabulary items. VTF has been employed to increase the familiarity of a writing framework by making it more similar to a mechanical writing framework [16]. Furthermore, studies have discovered that people are inherently skilled in identifying patterns of vibrotactile notes [21]. We hypothesize that the addition of VTF into the writing environment will improve recall of vocabulary items by enriching the encoding procedure with an additional input.

RQ4: How do the factors of annotation mode, presentation sequence and haptic feedback interact and which combination is most conducive for recall of vocabulary items in a mobile presentation framework?

H4: Ww.Seq.VTF will lead to the highest scores in immediate and delayed cued and free recall. Based on the rationale we have explained in the previous sections, we retained three factors that we hypothesize will affect the recall of vocabulary items. Their interaction will uncover the most conducive combination to present vocabulary. We hypothesize that the combination of whole word annotation, sequential presentation, and VTF will yield the best recall performance.

STUDY DESIGN

L2 language choice and word list

For the first language (L1), we chose English. For our choice of L2, we needed the following qualities. The words needed to be distributed across experimental conditions with uniform complexity and that none of the vocabulary items contained any linguistic cognates. Linguistic cognates are words having the same linguistic origin as one another (like, English father, German Vater, Latin pater). Linguistic cognates are easy to learn and less susceptible to forgetting than non-cognates [17]. To avoid any confounds introduced by items which are easier to learn than the rest, cognates had to be avoided.

We had to ascertain that it was a participants' first exposure to each of the vocabulary items. While we could ask about their familiarity with any language, there was no way to ensure that they had never been exposed to any of the L2 words without revealing the L2 words. Also, the words needed to be pronounceable by participants who were fluent in English. L2 words which were difficult to pronounce would add both variability and cognitive load.

Macedonia and Knösche [38] give a comprehensive overview of motivation behind choosing an artificial corpus. We adapted the same rationale and used the Wuggy pseudoword generator [30] to generate the pseudo corpus. We choose the Wuggy pseudoword generator as it facilitated the generation of polysyllabic pseudowords which followed English phonotactic constraints. Phonotactic constraints are rules which define what sound sequences are possible in a language. For example, gemination or consonant twinning which refers to lengthening the articulation of a consonant when compared to a single instance of the same type of consonant is not allowed in English, but is seen in many other languages such as French or Dutch. Another example is that letter combinations such as *ntat or *rkoop, are not allowed in English [27]

Word list

The word list defined the English words which would act as the meanings for the Wuggy generated pseudo vocabulary. These words needed to be simple and uniformly familiar. Carter [13] refers to Ogden's basic English as a core vocabulary which is designed to meet communicative adequacy. In other words, Ogden's vocabulary consisted of those words which are introduced early on for any speaker. We thus choose Ogden's list of 200 things for our L1 list. We used an 84 item vocabulary with L2 words generated using Wuggy pseudoword generator. Firstly, 200 pseudowords were generated and mapped randomly to the wordlist of Ogden's vocabulary. Then, 84 words were randomly chosen and divided into 6 groups of 14 ensuring uniformity of word length across the groups. Word

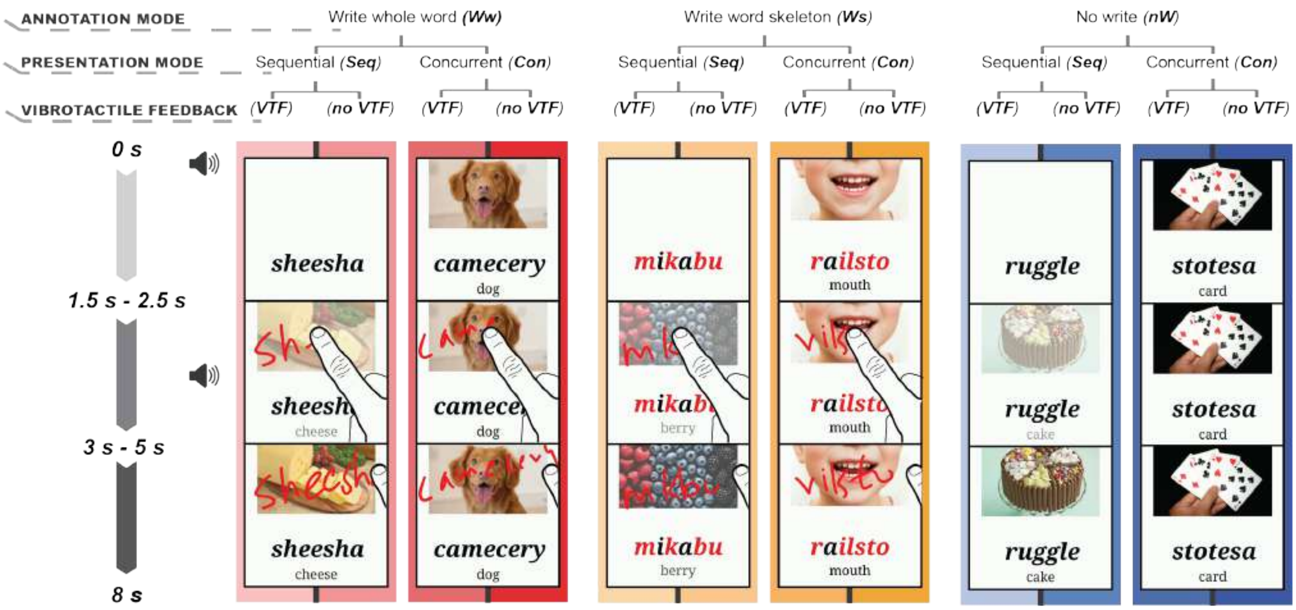

Figure 4. The different phases of the layout from 0s up to 8s (top to bottom) for the 12 conditions. Participants hear the audio pronunciation at 0s and once more between 1.5-5s based on writing for Ww and Ws, or at 4s for nW.

length was uniform as literature suggests that ease of recall is dependent on word length [3]. This assignment of vocabulary items to group was the same for all participants. We created audio pronunciations corresponding to the 84 selected L2 words using a Text to speech software. Each audio file was 1-2s. We choose images corresponding to the words from Google image search results, selecting the first result which met the clarity, relevance, and license to use criteria. All the audio files were in WAV format and images in JPEG/PNG format.

Layout elements

Audio: For all levels, audio played once at 0s. For No write level, audio played the second time at 4s. For write levels, strokes were monitored and audio played when strokes reached midscreen.

Annotation prompt: For write whole word level, participants were instructed to write the whole of L2 word. For word skeleton annotation, they were asked to write only the red letters in L2 word as shown in figure 4.

Progress to next word: For No write level, application proceeded to next word after 8s. For write levels, it proceeded 2.5 s after participants stopped writing.

VTF: For write levels, the vibration was active for as long as the participant was writing on the screen. For No write, the vibration was set to last for 4 s and was activated as soon as the application proceeded to the word.

Experimental method

Participants

We recruited 12 participants from the university community (5 female; M age=24.65, SD = 3.24). All participants were undergoing training in English medium university curriculum with self-rated English proficiency 8.34/10.0 (min 7). They were paid the standard compensation, an equivalent of 7.24 United States Dollars for their time.

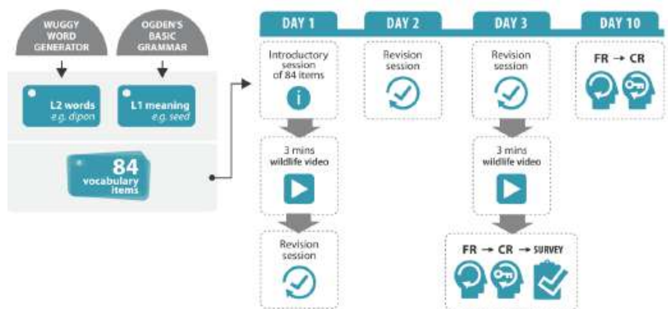

Figure 5. The complete experimental set-up from day 1 to day 10. FR: Free recall test, CR: Cued recall test.

Apparatus

All sessions were carried on a OnePlus 6 phone (1080 x 2280 pixels, dim. 155.7 x 75.4 x 7.8 mm, weight: 177g) in a quiet, closed room with only the experimenter and one participant present at a time. Participants sat on a chair with a desk of comfortable height in front of them with the phone in their non-dominant hand and wrote with their dominant index finger. They used earphones during the sessions. Assessment and post-questionnaire Google forms were filled on a standard laptop.

Sessions

Figure 5 illustrates the complete flow of the experiment. Each participant had 4 sessions in total for the 84 words spanned across 3 days. We decided on training for 3 days based on the implications of the results from Macedonia Knösche [38] which reports that the benefits of the experimental group with enactment encoding were evident from Day 3. In line with findings from Motor Neuroscience, they reason that motoric learning takes longer time to produce results as there is a need for consolidation processes after practice [4].

The initial session was designed to be common across all words and conditions. For this session, the 84 words were sorted by length, then alphabetically and presented to the participants sequentially. Participants were exposed to each word layout for 8s. They heard the audio twice, once at 1s and then once more at 4s.

The revision session started after a 3-min break with a relaxing video of wildlife. It presented the 84 items divided into 12 groups counterbalanced in order for the annotation mode and presentation sequence. Each combination of *Annotation mode X Presentation sequence* had 14 items. These 14 items were then randomly divided into 2 sets, one for each vibrotactile mode.

Participants were given clear instructions about the order of the modes and what, if at all they had to write. They were told their objective was to memorize as many pairs as possible. The second and third revision sessions were conducted 24 hours and 48 hours after the first revision sessions. They were given the immediate test Google form after revision session 3. They had no access to the materials except during the experiments. In addition to the standard compensation, they were told that every correct answer in their assessment would earn an equivalent of 5 US cents.

Testing

The test was divided into two sections, free recall followed by cued recall and recognition [59]. In free recall, participants were asked to list as many L1-L2 pairs as they could recall. They were not given any time limit. This was followed by cued test with 3 types of questions; cued L1 to L2 recall, cued L2 to L1 recall and cued L2 to L1 recognition. In cued English to L2, participants were given the English word and were asked to recall and write the corresponding L2 word in a text box provided. Similarly, they were given the L2 word and were asked to recall the English word in cued L2 to English. In the cued L2 to English recognition, they were given the L2 word and were asked to choose the corresponding English word from a set of four choices. Participants were told not to guess, however, they were instructed to write answers even in cases where they were unsure of the spellings.

The survey and interview was conducted immediately after the tests. The interviews were audio recorded. They were not informed about the 7-day delayed test which had the same free and cued assessment as the immediate test.

Scoring

Free recall section: In this section, every correctly written word-meaning pair received 1 point. If the spellings of the L2 words were incorrect, we calculated the error on a participant's answer using the Levenshtein [34] distance, which counts the minimum number of insertions, deletions and substitutions needed to correct the spelling and used the formula suggested by Ibrahim et al. [28]. We ignored any typos or spelling errors in the L1 word. (Example, if tongue is written as toungue) we still counted it as a correct answer as English word learning was not a focus of the training procedure.

Cued recall and recognition: All correct answers were given 1 point. The total assessment was for 84.0 points. In cued English to L2 recall, in case of spelling errors, we used similar scoring as free recall. In cued L2 to English recall and cued L2 to English recognition, correctly written answers, regardless of typos in L1, and correctly chosen options were given 1 point.

RESULTS

Cued recall and recognition

Immediate test results

The mean score in immediate cued recall test was 53.71 out of maximum possible score of 84.0 points and the standard deviation was 11.40. A repeated measures ANOVA was run on the immediate cued recall and recognition scores of the 12 participants to examine the effect of annotation mode, presentation sequence and vibrotactile feedback on the test performance. There was a significant effect of the annotation mode on the cued recall scores, ($F_{2,22}$ = 16.17, p < 0.001, η^2 = 0.223) indicating annotation completeness played a critical role in determining the benefits on recall. There was also a significant interaction between annotation mode and presentation sequence, ($F_{2,22}$ = 10.62, p < 0.001, η^2 = 0.061).

Post Hoc analysis with Bonferroni correction on the annotation mode was conducted and the mean differences and p_{bonf} are listed below. Mean difference(Ww, Ws) = 5.81, p_{bonf} = 0.003; Mean difference(Ww, nW) = 13.88, p_{bonf} = 0.001; Mean difference(Ws, nW) = 8.07, p_{bonf} = 0.054.

7-day delayed test results

The mean score in the 7-day delayed test was 50.21 out of maximum possible score of 84.0 points with standard deviation being 11.05. A repeated measures ANOVA was run on the 7-day delayed cued recall and recognition scores of the 12 participants to examine the effect of annotation mode, presentation sequence and vibrotactile feedback on the test performance. There was still a significant effect of annotation mode, ($F_{2,22}$ = 21.65, p < 0.001, η^2 = 0.271), however, the effect of the interaction between annotation mode and presentation sequence was no longer significant. Post Hoc analysis with Bonferroni correction on the annotation mode was conducted and the mean differences and p_{bonf} are listed below. Mean difference(Ww, Ws) = 5.77, p_{bonf}= 0.057; Mean difference(Ww, nW) = 14.87, p_{bonf}< 0.001; Mean difference(Ws, nW) = 9.1, p_{bonf}= 0.002.

Free recall

Immediate test results

The mean score in immediate free recall test was 21.87 out of maximum possible 84.0 and the standard deviation was 7.46. A repeated measures ANOVA was run and revealed that annotation mode showed significant effect ($F_{2,22}$ = 7.39, p = 0.004, η^2 = 0.089) and there was a significant effect of the vibrotactile feedback mode, ($F_{1,11}$ = 5.91, p = 0.033, η^2 = 0.027). Post Hoc analysis with Bonferroni correction on the annotation mode was conducted and the mean differences and p_{bonf} are listed below. Mean difference(Ww, Ws) =

2.21, p_{bonf}= 0.986, Mean difference(Ww, nW) = 8.41, p_{bonf}= 0.032; Mean difference(Ws, nW) = 5.93, p_{bonf}= 0.013. Also, Mean difference(VTF, nVTF) = 2.64, p_{bonf}= 0.078.

7-day delayed test results

The mean score in 7-day delayed free recall test was 15.67 out of maximum possible 84.0 and the standard deviation, 7.61. A repeated measures ANOVA was run and revealed that annotation mode showed significant effect ($F_{2,22}$=26.96, p <0.001, η^2 = 0.225). Post Hoc analysis with Bonferroni correction on the annotation mode was conducted and the mean differences and p_{bonf} are listed below. Mean difference(Ww, Ws) = 2.49, p_{bonf}= 0.220, Mean difference(Ww, nW) = 10.22, p_{bonf}< 0.001; Mean difference(Ws, nW) = 8.12, p_{bonf}< 0.001.

DISCUSSION

We have organized the discussion of this paper into parts answering the research questions. These can be used as references to each RQ in the flow of the whole discussion.

RQ1: How does the adding whole and partial word annotations to the revision phase of vocabulary learning affect recall? Through our hypothesis, H1.1, we had predicted a superior performance of the write levels of annotation mode. It is evident from our results that revising using annotations after being introduced to vocabulary items with flashcards enhances recall. We attribute the learning gains in the write conditions to the relevant motoric engagement achieved by annotating on the touchscreen. We therefore provide evidence to support our hypothesis H1.1 which predicted that recall of vocabulary items can be improved with annotations.

RQ2: What is the relationship between temporal sequence of presentation of various modal elements and recall of vocabulary items? Ww and Ws performed significantly better than nW in all recall tests. In the interviews, some participants were skeptical about the Ws condition. P4 remarked, *"I didn't like choosing the red letters, it sort of drew my attention away from actually learning the word"*. For Ww, participants were instructed to write the entire L2 word. However, for Ws, they need to pick out the letters corresponding to the word skeleton by choosing only the red letters. While this seems to be a trivial task, there was definitely an additional visual task involved. It is likely that the superiority of the sequential presentation is due to this additional visual task which adds to the cognitive load. Even though there was an additional visual task for Ws, our design used this approach to ensure uniformity of the number of times a learner was presented with the L2 word in each exposure. If the design had individual representations for the whole L2 word and the word skeleton, then the learner would be exposed to the L2 word twice, albeit in different forms in the Ws level. We acknowledge that this is an experiment-introduced limitation. For the whole word annotation, there was no such additional task. Participants were exposed to the complete layout with the image element, L1 and L2 verbal elements altogether in the concurrent presentation mode. This might have helped them with forming connections as described by Mayer [42] .

Further, this also explains why the interaction between the annotation mode and presentation sequence is significant in the immediate tests but not in the delayed test. Jones [29] conducted immediate and delayed tests of vocabulary after training with either pictorial, written, a combination of pictorial and written or no glosses to a French text for students who were learning French through English. The results of this study showed a steep decrease in the performance of the combined pictorial and written condition from immediate to the 3-week delayed cued recall test. They explain this result by attributing the good performance in immediate cued recall tests of the pictorial and written group to the richer and redundant amount of information, as is in our case with Ww. It reasons that the vast amount of information might have caused an overload of visual information that, with time, become "cluttered" and inhibited the ability to extract the required translation in the delayed test.

In our design, it is likely that there is a high level of element interaction [63] between the visual elements in the concurrent presentation level with Ww. However, for concurrent annotation with word skeleton writing, it is likely that participants used their visual attention to choose and copy the red letters, thereby eliminating the chance of acquiring all the visual information together. The delayed test performance in cued recall supports this reasoning as almost equal scores for Ww.Seq and Ww.Con were recorded. Whereas for word skeleton annotations, Ws.Seq performs consistently better in both immediate and delayed tests. These results provides partial evidence for the implication i3, suggesting that presentation in sequence counters the cognitive overload and is utilitarian in handling applications that need to present high quantities of information in the visual channel. Hypothesis H2.1 which predicted that sequential presentation would outperform concurrent presentation could therefore not be proven.

There is a fundamental difference in the retrieval mechanisms for free and cued recall. Tulving and Pearlstone [66] investigated this difference by distinguishing between the availability of a memory element and the ability to access it. Free recall is highly dependent on the accessibility of the memory elements. A learner that has nothing prompting them towards the memory element might not be able to access it even though it has not been lost. On the other hand, in cued recall, the retrieval conditions are different. Our results indicate that word annotations, both skeleton and whole, improve the accessibility of the vocabulary items thus perform equally well in free recall. However, whole word writing seems to have a clear advantage in encoding vocabulary items thereby making them available as shown by the results of cued recall.

On the other hand, both delayed and immediate cued recall tests saw a significant difference in Ww and Ws, indicating that the availability of vocabulary items is likely to benefit more from whole word annotations. Availability of any memory item depends on the trace formation. Our investigation, while preliminary, indicates that whole word annotation are more likely to help with trace formation of a memory item.

A thorough analysis in this direction could help target the most beneficial form of word annotation for any application depending on the intended outcome. If accessibility of memory item is the key objective, either form of annotation might be applicable. This can be very appropriate in language learning as most situations which demand retrieval of a vocabulary item inherently present with contextual cues. Macedonia and Knösche [38] highlight that learners acquire vocabulary items with a context and rarely as individual units.

However, in applications where the availability is the focus, such as memorization of elaborate procedures in, a recipe or a mathematical algorithm, which needs to be recalled without cues, whole word annotations are likely to benefit better.

RQ3: How does adding vibrotactile feedback (VTF) to writing affect recall of vocabulary? Prewett et al. [49] analyses 45 studies which investigate the efficiency of task performance under baseline and vibrotactile conditions. The analysis upholds vibrotactile cues as effective additions to a multi-modal presentation and suggests using VTF in addition to visual elements. Our results indicate that addition of VTF resulted in a significant improvement in immediate free recall. Enriching the material with the vibrotactile input showed a transient effect evident only in the immediate free recall. Given that VTF has significantly improved only free recall, we may assume that it helps with the accessibility of the vocabulary items. We may cautiously interpret these results using Sweller's cognitive load theory [63] which distinguishes between the working memory and long-term memory. It highlights that redundant multisensory information will only affect the working memory. This information becomes less volatile only when processed and organized into the long-term memory. It is likely that the VTF aids for encoding into the working memory but fails to exhibit the same benefits on the long-term memory. Therefore, even though the effects of VTF were transient, we propose that when integrated with techniques of spaced repetition [71], VTF might prove to be effective.

RQ4: How do the factors of annotation mode, presentation sequence and haptic feedback interact and which combination is most conducive for recall of vocabulary items in a mobile presentation framework? Even though our results did not see a significant three-way interaction between the three factors, we propose that each of these factors can be employed to improve specific elements of vocabulary recall. In summary, annotation mode was critical in determining the benefits of recall. Using whole word annotations recorded a mean score of 20.67 out of 28.0 in delayed cued recall over the 12.16 out of 28.0 for the No write conditions. Whole word annotations perform best when integrated with concurrent presentation.

INTEGRATION WITH EXISTING LANGUAGE PLATFORMS

Language learning is a comprehensive activity with multiple components such as vocabulary acquisition, semantics, phonology and grammar [40] and several existing mobile vocabulary apps cater to multiple of these components. Findings from our experiments can be integrated into existing mobile vocabulary apps, of which we provide examples with two applications below. We picked two of the top mobile vocabulary learning apps from android app store, Duolingo: Learn Languages Free(>100m downloads) and Learn Languages with Memrise(>10m downloads).

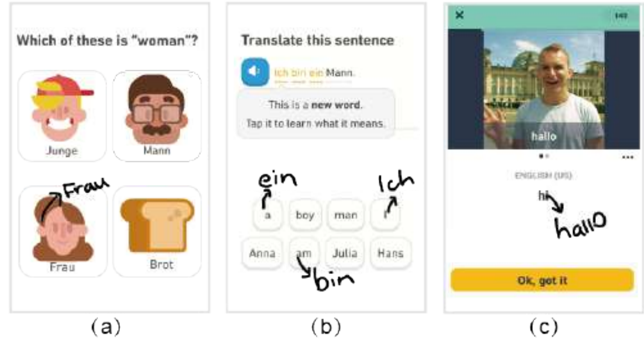

Figure 6. An illustration of a possible integration of our setting with Duolingo and Memrise. (a) and (b) layouts from Duolingo, show two activities in which learner has annotated. (c) from Memrise, shows a similar setting as our study.

In figure 6a, Duolingo has introduced the word *frau* and has prompted the learner with the English equivalent "woman" along with a clipart image of a woman. The learner can annotate the L2 word, *frau* next to the image of the woman. Figure 6b shows the second activity in which our feature can be incorporated in Duolingo.

As seen in figure 6c, the layout presents the L2 word with the L1 word and supplements it with video recording with presenter(s) speaking the word. The writing layout can be integrated into the application thereby allowing learners to benefit from writing when they are exposed to the L2 word.

LIMITATIONS

Although our studies reveal significant results, they are based on moderate sample size of participants. While we believe our results are valid, it can benefit from additional replication research to preserve and strengthen the scientific integrity of our findings [11]. Future studies can explore a broader demographic or age-based inclusion criterion. In addition, the ecological validity of the results can be enhanced with real and longitudinal deployment of vocabulary acquisition with haptically-enhanced mobile learning applications.

CONCLUSION

This study contributes the first investigative analysis of setting which achieves motoric engagement making use of haptic capabilities of touchscreen devices to improve vocabulary recall. We conducted a series of three pilots to identify design factors which affect recall of vocabulary items in a haptic-integrated setting. We recognized three design factors, annotation mode, presentation sequence, and vibrotactile feedback, that influence recall. We then evaluated these factors in a within-subject comparative study and supported our hypothesis that our setting significantly improves vocabulary learning. Our proposed setting is easily integrable into existing vocabulary applications and will benefit language learners to acquire vocabulary more easily.

ACKNOWLEDGEMENT

This work is supported in part by the NUS School of Computing Strategic Initiatives.

REFERENCES

[1] Shirley Agostinho, Sharon Tindall-Ford, Paul Ginns, Steven J Howard, Wayne Leahy, and Fred Paas. 2015. Giving learning a helping hand: finger tracing of temperature graphs on an iPad. *Educational Psychology Review* 27, 3 (2015), 427–443.

[2] Khalid Al Seghayer. 2001. The effect of multimedia annotation modes on L2 vocabulary acquisition: A comparative study. *Language Learning & Technology* 5, 1 (2001), 202–232.

[3] Alan D Baddeley, Neil Thomson, and Mary Buchanan. 1975. Word length and the structure of short-term memory. *Journal of verbal learning and verbal behavior* 14, 6 (1975), 575–589.

[4] Pierre Baraduc, Nicolas Lang, John C Rothwell, and Daniel M Wolpert. 2004. Consolidation of dynamic motor learning is not disrupted by rTMS of primary motor cortex. *Current Biology* 14, 3 (2004), 252–256.

[5] Joe Barcroft. 2006. Can writing a new word detract from learning it? More negative effects of forced output during vocabulary learning. *Second Language Research* 22, 4 (2006), 487–497.

[6] Joe Barcroft. 2007. Effects of word and fragment writing during L2 vocabulary learning. *Foreign Language Annals* 40, 4 (2007), 713–726.

[7] Lawrence W Barsalou. 1999. Perceptual symbol systems. *Behavioral and brain sciences* 22, 4 (1999), 577–660.

[8] Jennifer S Beaudin, Stephen S Intille, Emmanuel Munguia Tapia, Randy Rockinson, and Margaret E Morris. 2007. Context-sensitive microlearning of foreign language vocabulary on a mobile device. In *European conference on Ambient intelligence*. Springer, 55–72.

[9] Stephen Brewster and Lorna M Brown. 2004. Tactons: structured tactile messages for non-visual information display. In *Proceedings of the fifth conference on Australasian user interface-Volume 28*. Australian Computer Society, Inc., 15–23.

[10] Carrie J Cai, Philip J Guo, James Glass, and Robert C Miller. 2014. Wait-learning: leveraging conversational dead time for second language education. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2239–2244.

[11] Kelly Caine. 2016. Local Standards for Sample Size at CHI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 981–992. DOI: http://dx.doi.org/10.1145/2858036.2858498

[12] Manuel Carreiras, Margaret Gillon-Dowens, Marta Vergara, and Manuel Perea. 2008. Are vowels and consonants processed differently? Event-related potential evidence with a delayed letter paradigm. *Journal of Cognitive Neuroscience* 21, 2 (2008), 275–288.

[13] Ronald Carter. 2012. *Vocabulary: Applied linguistic perspectives*. Routledge.

[14] Chih-Ming Chen and Ching-Ju Chung. 2008. Personalized mobile English vocabulary learning system based on item response theory and learning memory cycle. *Computers & Education* 51, 2 (2008), 624–645.

[15] George M Chinnery. 2006. Emerging technologies: Going to the MALL: Mobile assisted language learning. *Language learning & technology* 10, 1 (2006), 9–16.

[16] Youngjun Cho, Andrea Bianchi, Nicolai Marquardt, and Nadia Bianchi-Berthouze. 2016. RealPen: Providing realism in handwriting tasks on touch surfaces using auditory-tactile feedback. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. ACM, 195–205.

[17] Annette MB De Groot and Rineke Keijzer. 2000. What is hard to learn is easy to forget: The roles of word concreteness, cognate status, and word frequency in foreign-language vocabulary learning and forgetting. *Language learning* 50, 1 (2000), 1–56.

[18] Adam K Dubé and Rhonda N McEwen. 2015. Do gestures matter? The implications of using touchscreen devices in mathematics instruction. *Learning and Instruction* 40 (2015), 89–98.

[19] New American Economy. 2017. Demand for Bilingual Workers More than Doubled in 5 years, New Report Shows. (2017). Retrieved September 20, 2019 from https://www.newamericaneconomy.org/press-release/demand-for-bilingual-workers-more-than-doubled-in-5-years-new-report-shows/.

[20] Darren Edge, Elly Searle, Kevin Chiu, Jing Zhao, and James A Landay. 2011. MicroMandarin: mobile language learning in context. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 3169–3178.

[21] Mario Enriquez and Karon MacLean. 2008. The role of choice in longitudinal recall of meaningful tactile signals. In *2008 Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*. IEEE, 49–56.

[22] Ram Frost and Shlomo Dentin. 1992. Reading consonants and guessing vowels: Visual word recognition in Hebrew orthography. In *Advances in psychology*. Vol. 94. Elsevier, 27–44.

[23] Nesta Futurelab, Laura Naismith, Peter Lonsdale, Giasemi Vavoula, Mike Sharples, and Nesta Futurelab Series. 2004. Literature review in mobile technologies and learning. (2004).

[24] Arthur M Glenberg, Tiana Gutierrez, Joel R Levin, Sandra Japuntich, and Michael P Kaschak. 2004. Activity and imagined activity can enhance young children's reading comprehension. *Journal of educational psychology* 96, 3 (2004), 424.

[25] Arthur M Glenberg and David A Robertson. 1999. Indexical understanding of instructions. *Discourse processes* 28, 1 (1999), 1–26.

[26] Joachim Grabowski. 2007. The writing superiority effect in the verbal recall of knowledge: Sources and determinants. *STUDIES IN WRITING* 20 (2007), 165.

[27] Heidi Harley. 2017. *English words: A linguistic introduction*. John Wiley & Sons.

[28] Adam Ibrahim, Brandon Huynh, Jonathan Downey, Tobias Höllerer, Dorothy Chun, and John O'Donovan. 2017. ARbis Pictus: a study of language learning with augmented reality. *arXiv preprint arXiv:1711.11243* (2017).

[29] Linda Jones. 2004. Testing L2 vocabulary recognition and recall using pictorial and written test items. *Language Learning & Technology* 8, 3 (2004), 122–143.

[30] Emmanuel Keuleers and Marc Brysbaert. 2010. Wuggy: A multilingual pseudoword generator. *Behavior research methods* 42, 3 (2010), 627–633.

[31] Zae Myung Kim, Suin Kim, Alice Oh, and Ho-Jin Choi. 2017. An adaptive vocabulary learning application through modeling learner's linguistic proficiency and interests. In *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)*. IEEE, 434–436.

[32] Dejan Kovachev, Yiwei Cao, Ralf Klamma, and Matthias Jarke. 2011. Learn-as-you-go: new ways of cloud-based micro-learning for the mobile web. In *International Conference on Web-Based Learning*. Springer, 51–61.

[33] Batia Laufer. 1989. A factor of difficulty in vocabulary learning: Deceptive transparency. *AILA review* 6, 1 (1989), 10–20.

[34] Vladimir I Levenshtein. 1966. Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*, Vol. 10. 707–710.

[35] Joel R Levin. 1981. On functions of pictures in prose. *Neuropsychological and cognitive processes in reading* 203 (1981).

[36] Chih-Cheng Lin and Ya-Chuan Yu. 2017. Effects of presentation modes on mobile-assisted vocabulary learning and cognitive load. *Interactive Learning Environments* 25, 4 (2017), 528–542.

[37] Robert W Lindeman, John L Sibert, and James K Hahn. 1999. Hand-held windows: towards effective 2D interaction in immersive virtual environments. In *Proceedings IEEE Virtual Reality (Cat. No. 99CB36316)*. IEEE, 205–212.

[38] Manuela Macedonia and Thomas R Knösche. 2011. Body in mind: How gestures empower foreign language learning. *Mind, Brain, and Education* 5, 4 (2011), 196–211.

[39] Manuela Macedonia and Katharina von Kriegstein. 2012. Gestures enhance foreign language learning. *Biolinguistics* 6, 3-4 (2012), 393–416.

[40] Brian Macwhinney. 2017. Language Acquisition The Basic Components of Human Language, Methods for Studying Language Acquisition, Phases in Language Development. (2017). Retrieved January 5, 2020 from `https://education.stateuniversity.com/pages/2153/Language-Acquisition.html`.

[41] Scott C Marley and Zsuzsanna Szabo. 2010. Improving children's listening comprehension with a manipulation strategy. *The journal of educational research* 103, 4 (2010), 227–238.

[42] Richard Mayer and Richard E Mayer. 2005. *The Cambridge handbook of multimedia learning*. Cambridge university press.

[43] Richard E Mayer and Roxana Moreno. 2003. Nine ways to reduce cognitive load in multimedia learning. *Educational psychologist* 38, 1 (2003), 43–52.

[44] Makiko Naka. 1998. Repeated writing facilitates children's memory for pseudocharacters and foreign letters. *Memory & cognition* 26, 4 (1998), 804–809.

[45] Mark Paterson. 2007. *The senses of touch: Haptics, affects and technologies*. Berg.

[46] Lyn Pemberton, Marcus Winter, and Sanaz Fallahkhair. 2010. Collaborative mobile knowledge sharing for language learners. *Journal of the Research Centre for Educational Technology* 6, 1 (2010), 144–148.

[47] Daniel Reschke Pires and others. 2018. L2 vocabulary instruction: an analysis of smartphone applications for english learning. (2018).

[48] Marc Prensky. 2005. What can you learn from a cell phone? Almost anything! *Innovate: Journal of Online Education* 1, 5 (2005).

[49] Matthew S Prewett, Linda R Elliott, Ashley G Walvoord, and Michael D Coovert. 2011. A meta-analysis of vibrotactile and visual information displays for improving task performance. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 42, 1 (2011), 123–132.

[50] Mei Pu and Zheng Zhong. 2018. Development of a Situational Interaction Game for Improving Preschool Children'Performance in English-Vocabulary Learning. In *Proceedings of the 2018 International Conference on Distance Education and Learning*. ACM, 88–92.

[51] Graham Rawlinson. 2007. The significance of letter position in word recognition. *IEEE Aerospace and Electronic Systems Magazine* 22, 1 (2007), 26–27.

[52] Claudia Repetto, Elisa Pedroli, and Manuela Macedonia. 2017. Enrichment effects of gestures and pictures on abstract words in a second language. *Frontiers in psychology* 8 (2017), 2136.

[53] Jerzy Rubach. 1996. Shortening and ambisyllabicity in English. *Phonology* 13, 2 (1996), 197–237.

[54] Claire N Rubman and Harriet Salatas Waters. 2000. A, B seeing: The role of constructive processes in children's comprehension monitoring. *Journal of Educational Psychology* 92, 3 (2000), 503.

[55] Nuria Sagarra and Matthew Alba. 2006. The key is in the keyword: L2 vocabulary learning methods with beginning learners of Spanish. *The modern language journal* 90, 2 (2006), 228–243.

[56] Eli Saltz and David Dixon. 1982. Let's pretend: The role of motoric imagery in memory for sentences and words. *Journal of Experimental Child Psychology* 34, 1 (1982), 77–92.

[57] Eli Saltz and Suzanne Donnenwerth-Nolan. 1981. Does motoric imagery facilitate memory for sentences? A selective interference test. *Journal of Verbal Learning and Verbal Behavior* 20, 3 (1981), 322–332.

[58] Jacobijn Sandberg, Marinus Maris, and Pepijn Hoogendoorn. 2014. The added value of a gaming context and intelligent adaptation for a mobile learning application for vocabulary learning. *Computers & Education* 76 (2014), 119–130.

[59] James P Schmidt, Tom N Tombaugh, and Penny Faulkner. 1992. Free-recall, cued-recall and recognition procedures with three verbal memory tests: Normative data from age 20 to 79. *Clinical neuropsychologist* 6, 2 (1992), 185–200.

[60] scientiamobile. 2017. Smartphone vs Tablet Orientation: Who's Using What? (2017). Retrieved September 20, 2019 from `https://www.scientiamobile.com/smartphone-vs-tablet-orientation-whos-using-what/`.

[61] Alexander Skulmowski and Günter Daniel Rey. 2018. Embodied learning: introducing a taxonomy based on bodily engagement and task integration. *Cognitive research: principles and implications* 3, 1 (2018), 6.

[62] Craig J Sutherland, Andrew Luxton-Reilly, and Beryl Plimmer. 2016. Freeform digital ink annotations in electronic documents: A systematic mapping study. *Computers & Graphics* 55 (2016), 1–20.

[63] John Sweller. 2011. Cognitive load theory. In *Psychology of learning and motivation*. Vol. 55. Elsevier, 37–76.

[64] Saeed Taki and Saeed Khazaei. 2011. Learning Vocabulary via Mobile Phone: Persian EFL Learners in Focus. *Journal of Language Teaching & Research* 2, 6 (2011).

[65] Medha Tare, Cynthia Chiong, Patricia Ganea, and Judy DeLoache. 2010. Less is more: How manipulative features affect children's learning from picture books. *Journal of Applied Developmental Psychology* 31, 5 (2010), 395–400.

[66] Endel Tulving and Zena Pearlstone. 1966. Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior* 5, 4 (1966), 381–391.

[67] Jeffrey D Wammes, Tanya R Jonker, and Myra A Fernandes. 2019. Drawing improves memory: The importance of multimodal encoding context. *Cognition* 191 (2019), 103955.

[68] Jeffrey D Wammes, Melissa E Meade, and Myra A Fernandes. 2016. The drawing effect: Evidence for reliable and robust memory benefits in free recall. *The Quarterly Journal of Experimental Psychology* 69, 9 (2016), 1752–1776.

[69] Jeffrey D Wammes, Melissa E Meade, and Myra A Fernandes. 2018. Creating a recollection-based memory through drawing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 44, 5 (2018), 734.

[70] Stuart Webb. 2008. The effects of context on incidental vocabulary learning. *Reading in a foreign language* 20, 2 (2008), 232–245.

[71] Gerlinde Weimer-Stuckmann. 2009. *Second language vocabulary acquisition: spacing and frequency of rehearsals*. Ph.D. Dissertation.

[72] Margaret Wilson. 2002. Six views of embodied cognition. *Psychonomic bulletin & review* 9, 4 (2002), 625–636.

[73] Kathryn T Wissman, Katherine A Rawson, and Mary A Pyc. 2012. How and when do students use flashcards? *Memory* 20, 6 (2012), 568–579.

[74] L-H Wong and C-K Looi. 2010. Vocabulary learning by mobile-assisted authentic content creation and social meaning-making: two case studies. *Journal of Computer Assisted Learning* 26, 5 (2010), 421–433.