VocaBura: A Method for Supporting Second Language Vocabulary Learning While Walking

ARI HAUTASAARI, Interfaculty Initiative in Information Studies, The University of Tokyo, Japan

TAKEO HAMADA, Interfaculty Initiative in Information Studies, The University of Tokyo, Japan

KUNTARO ISHIYAMA, Graduate School of Information Science and Technology, The University of Tokyo, Japan

SHOGO FUKUSHIMA, Interfaculty Initiative in Information Studies, The University of Tokyo, Japan and PRESTO, Japan Science and Technology Agency (JST), Japan

Learning a new language is difficult and time-consuming. Apart from dedicated classroom study, second language (L2) learners often lack opportunities to switch their attention to vocabulary learning over other daily routines. In this work, we propose a method that enables L2 learners to study new vocabulary items during their dead time, such as when commuting to school or work. We developed a smartphone application, VocaBura, which combines audio learning with location-relevant L1-L2 word pairs to allow users to discover new vocabulary items while walking past buildings, shops and other locations. Our evaluation results indicated that Japanese beginner level English learners were able to retain more vocabulary items with the proposed method compared to traditional audio-based study despite being less aware of L2 vocabulary acquisition having occurred. In our second study, we report on the level of English vocabulary coverage for L2 learning achievable with our proposed method. We discuss several design implications for educational technologies supporting second language learning.

CCS Concepts: • **Human-centered computing** → **Interaction techniques**; **Empirical studies in HCI**; *Usability testing*.

Additional Key Words and Phrases: Second language vocabulary learning, audio learning, walking, context awareness

ACM Reference Format:

Ari Hautasaari, Takeo Hamada, Kuntaro Ishiyama, and Shogo Fukushima. 2019. VocaBura: A Method for Supporting Second Language Vocabulary Learning While Walking. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 4, Article 135 (December 2019), 23 pages. <https://doi.org/10.1145/3369824>

1 INTRODUCTION

Learning a new language takes time, and often a life-time of effort. For new learners of a language, memorizing a dozen new words can take several hours of practice [16], while mastering a language may require several years of dedicated work¹. Furthermore, the more morphologically and phonetically dissimilar a second language (L2)

¹<http://web.archive.org/web/20071014005901/http://www.nvte.gov/lotw/months/november/learningExpectations.html>

Authors' addresses: Ari Hautasaari, a.hautasaari@iii.u-tokyo.ac.jp, Interfaculty Initiative in Information Studies, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8654, Japan; Takeo Hamada, htakeo@iii.u-tokyo.ac.jp, Interfaculty Initiative in Information Studies, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8654, Japan; Kuntaro Ishiyama, kuntaro@nae-lab.org, Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8654, Japan; Shogo Fukushima, shogo@nae-lab.org, Interfaculty Initiative in Information Studies, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8654, Japan, PRESTO, Japan Science and Technology Agency (JST), 4-1-8 Honcho, Kawaguchi, Saitama, 332-0012, Japan.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.

2474-9567/2019/12-ART135 \$15.00

<https://doi.org/10.1145/3369824>

is to the learner's native language (e.g., Japanese and English), the more time investment is needed to reach proficiency [5, 8, 16].

In an effort to overcome these challenges, recent research on second language educational technologies have taken an approach to support microlearning (e.g., [8, 12, 34]). Microlearning connotes short and sporadic opportunities during the day when a learner shifts their focus from other tasks to vocabulary learning [4, 8, 12], such as studying by looking at a smartphone screen while waiting a train [4, 8]. Contextual microlearning approaches have further enriched traditional flashcard microlearning with contextually relevant information derived from nearby locations based on the learner's GPS coordinates [4, 8]. However, while approaches such as Edge et al. (2011) show promise and allow for new types of learning methodology, they still require the L2 learner to dedicate their time, albeit in short bursts, and shift their focus fully to the vocabulary learning activity.

Our goal is to support L2 vocabulary learning during daily routines, that is, while a user is focused on other tasks unrelated to vocabulary learning. In this work we introduce VocaBura (Figure 1), a context-aware audio-based vocabulary learning system, which enables L2 vocabulary learning during users' dead time [4, 8], such as while walking to and back from work or school. In our approach, rather than the learner reading through L2 words from a dictionary or a smartphone flashcard application, they recall words related to real-life locations they pass during their walking commute. In other words, vocabulary learning occurs when L2 learners *discover* new vocabulary items by accident during another daily activity unrelated to L2 study.

Fig. 1. Japanese participants learning the English word *horticulture* with the VocaBura application when walking past a garden.

As a user is walking, the VocaBura application tracks their GPS coordinates in relation to locations along the route, such as convenience stores, restaurants and parks. The location information is used to generate word pairs in the user's native language (L1) and the L2 they are studying based on the information related to a location and location category (e.g., garden) available in online services such as Google Maps and Wikipedia. In predetermined intervals, the user will hear an audio of a L1-L2 word pair *by accident* when passing a location, and by doing so discovers a new L2 word that is relevant to the real-world location (Figure 1).

The design of the VocaBura learning method follows the cognitive theory of multimedia learning as well as dual-coding theory, where we provide both visual (i.e., the appearance of a location associated with a word) and auditory cues for each L2 vocabulary item [23, 24]. Secondly, VocaBura allows for encoding specificity and situated cognition during L2 vocabulary learning by linking each L1-L2 word pair to the context of use [2, 8, 33], that is, a real-world location where a word may be used in practice. In this paper, we first focus on evaluating the efficacy of the VocaBura approach on L2 vocabulary recall. Secondly, we explore the potential coverage of English language words for L2 vocabulary learning with a fully automated VocaBura learning method.

In the following sections, we first discuss the related works and introduce our hypotheses and research questions. Next, we describe a controlled experiment with beginner level English language learners in which we compare the proposed VocaBura learning method with an audio-based L2 learning method, where the participants are fully engaged in the vocabulary learning task. Our experimental results suggest that beginner level English learners are able to retain significantly more English language words with the proposed method compared to the baseline audio-only learning method. Furthermore, we show that the participants were less aware that L2 vocabulary acquisition occurred when using the VocaBura method suggesting that they discovered at least some of the L2 words by accident while focusing on walking along a route between predetermined locations. Regarding our second study, we report on the number and difficulty level of English language words that can be studied using an automated VocaBura learning method, and by doing so demonstrate the efficacy of our approach for beginner and intermediate level L2 English learners. We further report on a case study, where we describe the potential L2 vocabulary coverage of the VocaBura learning method on two real-life walking routes. Lastly, we discuss how these results build upon and contribute to L2 microlearning, CTML and dual-coding theories, and draw practical implications for the design of novel second language learning technologies.

2 BACKGROUND AND RELATED WORK

The cognitive theory of multimedia learning (CTML) builds upon the assumptions that humans (or learners) have two separate channels for processing information; visual and auditory [23]. Secondly, learners have limited cognitive resources to process information in each channel at one time. Furthermore, learners actively process incoming information in to mental representations, and integrate these representations with other existing knowledge [23]. That is, CTML posits that learners do not interpret visual and auditory information in a mutually exclusive way, but rather dynamically organize this information to logical mental constructs as posited by the dual-coding theory [23, 24].

Dual-coding theory is a cognitive theory which essentially suggests that information is represented as both visual and verbal concepts in the human mind [27]. While the visual and verbal information are stored in different channels, humans can make use of either type of information to recall both the visual and the verbal information separately or simultaneously. For example, when asked to recall the word 'apple', humans are able to also conjure the image of an apple stored in their mind. Conversely, when presented with an image of an apple, they can simultaneously conjure the verbal information representing the word 'apple'.

In relation to L2 vocabulary learning, Paivio and Desrochers (1980) proposed a version of the dual-coding theory for bilinguals. This model suggests that L1 and L2 concepts are interdependent and stored in a shared-system with the visual information [27, 28]. In other words, bilinguals are able to recall the related visual information based on either the L1 or the L2 word, and conversely, can recall both the L1 and the L2 words based on the same visual information. The third vector in this interdependent system is recalling a L2 word based on either L1 verbal information, the shared visual information or a combination of both [28]. The expectation, thus, in L2 learning approaches based on CTML is that supporting the dual-coding of visual information with L1-L2 verbal information during a L2 vocabulary learning phase may improve the recall of L2 words on a later date as the

learners are able to recall the L2 words based not only on the L1 verbal information but also on the shared visual information [27, 28].

Previous research in second language acquisition (SLA) have put these assumptions to test by combining visual (e.g., pictures) and auditory (e.g., phonology of written words) cues in L2 vocabulary learning (e.g., [3, 29, 37]). Results from these previous works indicate that L2 learners recognize more new vocabulary items in post-tests when visual cues are presented alongside written text [3, 23, 24, 29, 37]. That is, by increasing the number of modalities in which learning material is presented, such as when combining textual captions with video [11], L2 learners are able to retain more of the studied L2 words in their long-term memory [37].

2.1 Second Language Microlearning

One of the oldest language learning methods, Rosetta Stone (1992), applies the principles described by CTML and dual-coding theories [23, 24]. The method combines audio and visual cues to help learners map mental representations of new L2 words without giving the L1 translations. Together with the audio-based Pimsleur system (1967), which predates CTML, these methods comprise the most popular lesson-based approaches. However, content in these lessons are typically provided in 20 to 30 minute chunks, which presents a challenge especially for learners who struggle with balancing the time investment between L2 study and other daily tasks [8, 13].

Recent solutions to overcome problems with time management, as well as motivation to keep up daily vocabulary study, have focused on microlearning [12]. The concept of microlearning involves utilizing moments during a learner's day when they can momentarily shift their focus fully to a learning task [8, 12], such as while waiting for a bus or a train. That is, instead of learners devoting 20 to 30 minutes a day to L2 study, microlearning approaches promote multiple short study sessions throughout the whole day.

Arguably, the first L2 microlearning method was the utilization of paper-based flashcards providing L1-L2 word pairs as the correct answer and the stimulus. The traditional flashcard approach has been since implemented in numerous software applications, often to automate the spaced repetition process of flashcard review derived from the Leitner system [13]. Such examples include Anki (2008), which allows users to create their own flashcard decks, or use pre-existing ones, and study them on either a PC or a smartphone interface [9].

With the increase in computing power and portability of smartphones, recent L2 microlearning applications have added further modalities to the traditional flashcard learning method. For example, applications have combined audio-based learning with mobile flashcards, such as MemReflex [7] which supports text-based as well as audio-based flashcard review and quick succession retesting. Tip Tap Tones [6], on the other hand, supports microlearning of Mandarin Chinese tonal sounds on a smartphone.

Besides combining visual and auditory information in L2 vocabulary learning, recent smartphone applications have also leveraged users' GPS location data to provide additional contextual information to L1-L2 flashcards. Situated cognition and encoding specificity in L2 vocabulary learning mean that words cannot be fully divorced from contexts, and are recalled the best when they are learned in the context of use (e.g., learning the word horticulture in a garden) [2, 8, 33]. MicroMandarin [8] supports contextual microlearning by combining contextually relevant content derived from nearby locations (e.g., shops, restaurants) based on users' GPS coordinates with spaced repetition of textual flashcards [13]. Similarly, Vocabulary Wallpaper [4] presents contextually relevant L1-L2 flashcards on learners' smartphone wallpaper instead of a dedicated L2 microlearning application to promote implicit acquisition of L2 vocabulary (i.e., by users accidentally glancing at the word pairs on their wallpaper).

However, while many of the studies on recent microlearning applications have reported promising results on L2 vocabulary acquisition, they still require users to shift their attention fully to the language learning activity in brief bursts. For example, Dearman and Truong (2012) reported that the users of Vocabulary Wallpaper focused on the L1-L2 flashcards on their smartphone wallpaper during their dead time, such as when waiting in line [4].

That is, the learners intentionally shifted their focus to the smartphone screen when they had a moment of free time as opposed to learning the L2 vocabulary while doing other tasks with the smartphone.

2.2 Intentional, Explicit, Incidental and Implicit Vocabulary Learning

Second language vocabulary learning can occur in a multitude of situations ranging from formal textbook study to accidentally picking up new vocabulary when overhearing a L2 conversation. Intentional learning connotes activities where the learner's goal is to memorize lexical information, such as when a learner aims to commit a L2 word list to memory before taking a test [16]. Generally speaking, this type of learning is also explicit, meaning that the learner is conscious about learning occurring through hypothesis testing, for example, by iteratively checking their own progress [32].

Incidental L2 vocabulary learning can be described by the perspectives in Krashen's Monitor Model [19–21]. Firstly, incidental vocabulary acquisition process is subconscious. That is, the learner is not aware that L2 vocabulary acquisition has occurred. Secondly, incidental L2 vocabulary acquisition requires minimum amount of cognitive effort. The third perspective in the Monitor Model states that incidental vocabulary acquisition occurs in a language environment where the L2 is prominent. In this study, we mainly focus on the first perspective; subconscious learning.

Much of the research conducted on incidental L2 vocabulary learning has focused on prolonged reading tasks, that is, incidental vocabulary learning through reading a book or a text passage [30]. Experimental results investigating L2 reading suggest that learners are able to recall more words they learned incidentally rather than intentionally (e.g., [1, 15]), and audio support (i.e., narration) during L2 reading can enhance incidental vocabulary learning [22]. One study comparing reading and listening for incidental L2 vocabulary learning reported that while reading outperformed listening for incidental vocabulary acquisition, listening was beneficial for long-term vocabulary retention in advanced L2 learners [35].

However, while incidental learning occurs non-explicitly (i.e., subconscious learning), the term connotes a conceptually different process to *implicit* learning [31]. Implicit learning is a psychological concept that means the absence of consciousness, whereas incidental learning is a pedagogical term that roughly means the lack of intention. The difficulty in differentiating between implicit learning and incidental learning stems from the unclear definition of the term *consciousness* [31]. In the context of L2 vocabulary learning, the question leading to the most terminological confusion seems to be whether unintentional learning can or cannot occur with consciousness.

Indeed, the definition of what constitutes as incidental learning, as opposed to intentional learning, has been through much debate in multiple academic disciplines including psychology and second language acquisition (SLA) (see, for example, [17]). Despite this, as Hulstijn (2008) argues, the term *incidental learning* is not likely to gain strong theoretical meaning in the SLA field due to difficulties in defining what actually constitutes as incidental instead of intentional learning. For example, if a person reads a foreign language book for the purposes of increasing their L2 vocabulary and happens to pick up a number of new second language words, at which point can this vocabulary acquisition be considered incidental?

Hence, it is important to note that in our work we consider *incidental learning* as a more informal concept connoting learning that occurs non-explicitly while learners are focusing on other tasks unrelated to vocabulary learning [16, 17, 32]. However, in order to avoid terminological confusion between the multitude of research fields that examine this phenomenon we opt to use *learning that occurs by accident* in the present work. Our aim is to explore to what extent L2 vocabulary learning with the proposed VocaBura learning method occurs non-explicitly, or conversely, to what extent are users aware of vocabulary learning having occurred while they were focused on another task unrelated to L2 study - walking along a specified route.

2.3 Hypotheses and Research Questions

Firstly, we are interested in exploring how well beginner level English learners are able to learn and retain L2 vocabulary items with our proposed method compared to a traditional audio-learning method. We hypothesized that visual (i.e., appearance of a location) [4, 29, 30, 37] and contextual cues (i.e., context of use) [2, 8, 33] associated with audio-based verbal information during language learning would improve the recall and retention of L2 words for beginner level L2 learners:

H1a: Beginner level L2 learners of English are able to perform better in an immediate vocabulary recall test with the proposed method compared to an audio-only learning method.

H1b: Beginner level L2 learners of English are able to perform better in a delayed vocabulary recall test with the proposed method compared to an audio-only learning method.

Next, since VocaBura users will be focusing on an unrelated task (i.e., walking along a route), we want to know whether and to what extent vocabulary learning with our proposed method might occur by accident. That is, to what extent are learners aware of L2 vocabulary acquisition having occurred with our proposed method [19–21]. Hence, we asked the following research question:

RQ1: Are beginner level L2 learners more or less aware of vocabulary acquisition with the proposed method compared to an audio-only learning method?

Lastly, we explore the level of English vocabulary coverage for L2 learning achievable with an automated VocaBura learning method in our second study. In Study 2, we ask the following two research questions:

RQ2: How many English words can L2 learners study with our proposed automated method?

RQ3: What difficulty level are the words L2 learners are able to study with our proposed automated method?

3 STUDY 1: METHOD

3.1 Overview

We conducted a laboratory experiment with a single factor (vocabulary learning method: VocaBura vs. baseline) within-subjects design. Seventeen native Japanese speakers participated in this experiment. They were asked to study 10 English vocabulary items with our proposed learning method (VocaBura) and 10 English vocabulary items with an audio-only learning method (baseline). We collected objective and subjective measures to assess the efficacy of our proposed learning method for L2 vocabulary recall (immediate and delayed), and participants' awareness of L2 vocabulary acquisition during the learning tasks.

3.2 Participants

Seventeen² native Japanese speakers participated in this experiment (11 female, 6 male). They were hired and compensated through a temp agency based on their self-reported *low English language skill*. Two of the participants had taken a standardized English language test (Test of English for International Communication, TOEIC), and reported scoring 400–450 out of 990 points. Fifteen of the participants had never taken a standardized English language test. The participants evaluated their English listening and reading proficiency on two 5-point scales (1=Not proficient at all, 5=Very proficient). Their average listening ($M=1.56$, $SD=0.51$) and reading ($M=1.50$, $SD=0.51$) self-evaluations indicated that they were not proficient in English (i.e., beginner level English learners). Their mean age was 35.00 ($SD=6.08$) years.

The reason why we focus on beginner level English language learners in Study 1 is two-fold. Firstly, beginner level language learners take considerably longer time to memorize even a small set of new L2 words compared to intermediate or advanced level learners [16]. Hence, we attest that beginner level learners as a user group would benefit from the VocaBura learning method the most. Secondly, we wanted to use a naturalistic approach in the

²We initially hired 20 native Japanese speakers, but excluded two participants for their non-compliance to the experiment protocol (risk of Type I error), and one participant for their high score in the pre-experiment vocabulary test (see: 5.1 Manipulation Checks).

controlled experiment, where the learning corpus is created based on real information available online. That is, advanced learners may already have a large L2 vocabulary making it challenging to control for the difficulty of L2 words and participants' pre-existing knowledge without using a pseudo language.

3.3 Learning Corpus

We created a Japanese (L1)-English (L2) location-based audio learning corpus consisting of 20 word pairs (Figure 2). For the purposes of this study, we chose five locations and shops along a simple walking route that were distinguishable regardless of the time of the day within a university campus. We then extracted the location name, category (e.g., garden, convenience store) as well as any direct link web page from Google Maps³, and searched for closest match Wikipedia article in Japanese⁴. Next, we used GensenWeb⁵ to extract a ranked keyword list based on the textual content in the Wikipedia articles and direct link web pages [25]. The semi-automatically created keyword list was first cleaned by omitting proper nouns, compound nouns and Wikipedia system terms, and then translated to English using Google Translate⁶.

Fig. 2. Semi-automated L1-L2 word pair generation process.

In order to control the difficulty of the Japanese-English word pairs, we filtered out all beginner level English words that the participants might be familiar with using the difficulty ratings in Weblio Japanese-English dictionary⁷. Weblio includes a numerical score between 1 (easy) and 16 (difficult) indicating the difficulty level of each English language word within its database⁸. The corpus was fully randomized to two sets of 10 L1-L2 word

³<http://maps.google.com/>

⁴<https://ja.wikipedia.org/>

⁵<http://gensen.dl.itc.u-tokyo.ac.jp/>

⁶<http://translate.google.com/>

⁷https://help.weblio.jp/ejje/endictionary/e_info/e_i_dictionary/h0116/ (in Japanese)

⁸<https://uw1.weblio.jp/info/about-vocab-test/> (in Japanese)

pairs for each participant (10 word pairs for the VocaBura method and 10 word pairs for the baseline method per participant). We chose the number of words in each set to match that of previous works with similar learning tasks, where the size of L2 learning corpora have generally ranged from 8 to 15 words [18, 26, 36].

Lastly, all Japanese and English word pairs were transformed to mp3 audio using Sound of Text text-to-speech software with a female voice⁹. Each participant was assigned two random sets which they would study with the proposed VocaBura method and the baseline learning method, respectively. The average difficulty level for the VocaBura learning corpus was 7.64 (SD=0.80, Range=2.4), and the average difficulty level for the baseline corpus was 7.96 (SD=0.80, Range=2.4). Results from a two-tailed t-test indicated that there was no significant difference in the English vocabulary difficulty between the VocaBura and baseline corpora ($t[33]=1.20$, $p=n.s.$).

3.4 VocaBura Application and Equipment

In this work, we asked the participants to use a stand-alone version of our proposed L2 vocabulary learning mobile application, VocaBura. The application tracks a user's GPS coordinates in real-time as they are walking past various locations. When the user arrives within a 25-meter radius from a target location, the VocaBura application will play an audio of a location-related L1-L2 word pair created by the process described in Chapter 3.3. The 25-meter radius was determined as a cutoff point based on results from a preliminary experiment to balance inaccuracies in GPS tracking, visibility of the target location from the user's point of view, and users' average walking speed [10, 14]. More specifically, in our initial prototype we set the radius at 30 meters for our preliminary experiments, but this lead the audio of word pairs to occasionally play before the target location was visible, especially for participants with slower walking speeds.

The VocaBura application was installed in a Nexus 6 Android smartphone with GPS functionality. The VocaBura user interface included a check sum on whether an audio file was played in the correct location for monitoring purposes and no other textual or visual content. In the baseline condition, the participants were given an iPad tablet. The baseline audio learning interface included an indicator signifying when 90 seconds had passed, which prompted the participants to listen to the next audio file. The 90 second indicator roughly matched the time it took participants to walk from one location to another in the VocaBura condition. The user interface included no textual content. The participants were given non-noise cancelling outer-ear headphones to listen to the L1-L2 audio.

In both conditions, the participants first heard an audio cue of three consecutive beeps before the Japanese-English audio as a notification that the next L1-L2 word pair was approaching. Audio of the L1-L2 word pairs were played three times with approximately one second pause between repetitions. These design choices were made based on feedback from a preliminary experiment [10, 14]. Participants in our preliminary experiment suggested adding an indicator that would inform them when a location related to a word was approaching to help them memorize the L2 words, which they had only heard once [14]. Hence, in addition to shortening the radius around each location, we added the audio-based notification sound, and increased the number of times a L1-L2 word pair audio was played from our initial VocaBura prototype.

3.5 Task and Procedures

All participants took a written pre-experiment vocabulary test in which they were presented with the 20 English words they were going to study during the experiment in randomized order. The participants were asked to translate any English words they knew to Japanese or leave their answer blank if they were unfamiliar with the word. This test was to ensure that none of the participants had previous knowledge of the second language words to be studied during the experiment. After filling the pre-experiment vocabulary test, they were given instructions on the L2 vocabulary learning tasks. In the VocaBura condition, the participants were told that their

⁹<https://soundoftext.com/>

task was to walk a predetermined route within a university campus, and that during the walk they would hear an audio of L1-L2 word pairs in the vicinity of certain locations (Figure 3). In total, the participants passed by five locations twice (forth and back), and heard an audio of different L1-L2 word pair each time (10 word pairs). The VocaBura learning task took a total of 15 minutes to complete.

Fig. 3. Designated walking route during the L2 vocabulary learning task with VocaBura (black circles represent a 25-meter radius around designated locations in which the audio of L1-L2 word pairs is played).

The participants completed the audio-only baseline condition on a different day one week after the VocaBura condition. In the baseline condition, the participants were asked to move to an outside public space within a university campus and either sit or stand while studying the L1-L2 word pairs. In the baseline condition, the participants listened to 10 L1-L2 word pairs in similar audio as in the VocaBura condition (three beeps and three repetitions) in 90 second intervals. The baseline learning task took a total of 15 minutes to complete.

After each learning session (10 word pairs), the participants were asked to return to the testing area and turn in their equipment. Once the participants were seated, they were asked to take a vocabulary recall test (immediate test), where they listened an audio of each English word in randomized order and wrote down a translation to the word in Japanese. The delay of approximately five minutes between the learning task and testing also served a function of minimizing the recency effect in the vocabulary recall test. After finishing the recall test, the participants filled an online survey about their learning experience.

After one week (7 days) the participants took a delayed recall test, where they listened to the same English words as in the immediate test in a randomized order and translated them to Japanese. They were not given any feedback on their vocabulary recall performance after neither the immediate nor the delayed test.

4 STUDY 1: MEASURES

We collected four types of measures: manipulation checks to ensure participants had no pre-existing knowledge of the L2 English words in the experiment corpus and that they heard the assigned words in their assigned locations, objective L2 vocabulary recall (immediate and delayed) to assess participants' learning performance, and a survey item measuring the participants' perceptions on L2 vocabulary acquisition.

Manipulation checks. The participants' pre-existing knowledge of the L2 English words was assessed in a pre-experiment written vocabulary test where they were asked to translate 20 English words to Japanese. Secondly, we gathered automated log data on whether the VocaBura software played the audio of L1-L2 word pairs in their assigned locations. In addition, we asked each participant whether they missed any of the words, for example, due to hardware malfunction.

Second language vocabulary recall. The participants' L2 vocabulary recall was measured in two tests; immediate and delayed. In both tests, the participants were asked to listen to an audio of the English words they studied during the learning task and write down the Japanese translation of each word. The immediate recall test was administered approximately five minutes after the learning task finished. The delayed recall test was administered after one week (7 days) for each experimental condition. In all tests, the order of the English words was fully randomized between the learning session, the immediate test and the delayed test. The participants score (0-10) was used as a measure of second language vocabulary recall.

Awareness about vocabulary acquisition. The participants' perceptions about their vocabulary acquisition was measured by a post-experiment questionnaire item which asked *"how many new English words do you think you learned using this system?"*

5 STUDY 1: RESULTS

5.1 Manipulation Checks

Seventeen participants scored 0 points out of 20 in the pre-experiment vocabulary test. One participant had pre-existing knowledge of the L2 English words in our experiment corpus (score of 4 out of 20), and their data was omitted from the following analyses. Automatically gathered log data and participants' self-reports indicated that all participants heard the audio of 10 Japanese-English word pairs at their designated locations in the VocaBura condition, and 10 word pairs in the baseline condition (100 percent).

5.2 Second Language Vocabulary Recall

Our H1a and H1b predicted that beginner level English language learners would perform better in the immediate and delayed L2 vocabulary recall tests, respectively, when they used the VocaBura learning method compared to an audio-only baseline method. To test our hypotheses, we compared the participants' average scores (0-10) in the immediate and delayed recall tests between the two learning methods. Secondly, we calculated the number of unlearned words for each participant (i.e., forgetting rate) and compared the average number of unlearned words between the two learning methods (Figure 4).

Results from a paired samples t-test indicated that the participants' mean score in the immediate vocabulary recall test was not significantly different between the VocaBura learning method ($M=4.41$, $SD=2.40$) and the baseline method ($M=5.18$, $SD=2.51$): $t[16]=1.20$, $p=n.s.$. However, the participants' mean test score in the delayed vocabulary recall test was significantly higher with the VocaBura method ($M=3.12$, $SD=1.83$) compared to the baseline method ($M=2.18$, $SD=1.70$): $t[16]=3.77$, $p<.01$. Furthermore, results from a paired samples t-test indicated that the participants unlearned significantly fewer words with the VocaBura learning method ($M=1.29$, $SD=1.45$) compared to the baseline method ($M=3.00$, $SD=2.37$) during the one week period between the immediate and delayed recall tests: $t[16]=2.60$, $p<.05$.

Fig. 4. Mean immediate and delayed L2 vocabulary recall test scores (left), and the mean number of unlearned words (right) (error bars represent the standard error of the mean).

These results indicated that contrary to H1a, beginner level English learners' performance was not significantly different in the immediate recall test when comparing the two learning methods. However, consistent with H1b, the participants scored significantly higher in the delayed recall test, and unlearned (i.e., forgot) significantly fewer English words with the proposed VocaBura method compared to the baseline learning method.

5.3 Awareness about Vocabulary Acquisition

In our RQ1, we asked whether beginner level English language learners would be more or less aware of L2 vocabulary acquisition with the proposed VocaBura learning method compared to an audio-only baseline method. Results from a paired samples t-test indicated that on average the participants estimated that they learned significantly fewer English words with the VocaBura method ($M=2.82$, $SD=2.21$) compared to the baseline method ($M=4.65$, $SD=2.45$): $t[16]=2.42$, $p<.05$. This result partially answered our RQ1. The participants were less aware that vocabulary acquisition had occurred with the proposed VocaBura learning method compared to the baseline method.

6 STUDY 2: L2 ENGLISH VOCABULARY COVERAGE WITH THE VOCABURA LEARNING METHOD

In Study 1, we investigated the level of L2 vocabulary recall with our proposed learning method. For the purposes of Study 1, we prepared a learning corpus specifically targeted at the participant demographic (i.e., Japanese beginner level English language learners) via a semi-automated process including a human review (Figure 2). In Study 2, we shift the focus on the generalizability of a fully automated VocaBura system, and explore the potential L2 vocabulary coverage with our proposed learning method.

Figure 5 illustrates the automated process for generating a location-related word list based on a location category. For the following analyses, we use all Google My Business¹⁰ categories listed before February 2018¹¹. Google My Business categories reflect all possible location categories available in Google Maps, including public services such as parks and schools, which are editable by Google users and hence subject to changes, omissions and additions over time. We searched for a matching article in the English Wikipedia for each of the categories

¹⁰<https://www.google.com/business/>

¹¹<https://classybrain.com/google-my-business-categories-2018/>

Fig. 5. Automated process for creating a location-related English word list for second language vocabulary learning with VocaBura.

using a full-text search algorithm (CirrusSearch¹²), which produced a Wikipedia page related to 3982 of the location categories. We then automatically extracted the textual content (i.e., sentences) within each article, and generated a keyword list for each location category by inputting all extracted text to Gensen Web keyword extractor¹³. By this process we were able to automatically generate an individual English word list related to each of the 3982 location categories. It is important to note that while each word list is generated based on the information available in Wikipedia for each Google My Business category, our method does not presently consider the level of relatedness of individual words within these lists to a specific location, and this aspect of our approach is part of our future work. Secondly, the reason for focusing on English in this study is two-fold. English Wikipedia is currently the largest Wikipedia language version available, and offers the potential for creating the largest learning corpus. Secondly, English is a popular second language studied in Asia, and thus using English as the target language caters to the initial target user group for VocaBura (see also: Study 1).

6.1 Vocabulary Coverage for L2 Learning

We used two types of English language vocabulary corpora to explore the potential coverage of our approach. The ALC¹⁴ Standard Vocabulary Level (SVL) list is aimed at Japanese English language learners, and includes 12,000 English words divided to 12 difficulty levels. In the ALC SVL list, levels 1 to 5 (i.e., the first 5000 words) include the vocabulary needed to pass the English language test in a Japanese university entrance examination, whereas the levels 1 to 10 (i.e., the first 10,000 words) include the English vocabulary needed to gain the highest score in TOEIC standardized English language test. Levels 11 and 12 include advanced vocabulary considered to be essential for fluent conversation with native speakers¹⁵.

¹²<https://www.mediawiki.org/wiki/Extension:CirrusSearch>

¹³<http://gensen.dl.itc.u-tokyo.ac.jp/>

¹⁴<https://www.alc.co.jp/> (in Japanese)

¹⁵<https://www.alc.co.jp/vocogram/article/svl/> (in Japanese)

The second corpus is comprised of five individual corpora from the New General Service List¹⁶. We included the Core General English Words (NGSL) list including 2800 words, Core Spoken (NGSL-S) list with 700 words, Academic Words (NAWL) list with 960 words, TOEIC Words (TSL) list with 1200 words, and Business Words (BSL) list with 1700 words in our analyses. Combined, these corpora include 7360 English language words that cover approximately 92 percent of vocabulary used in English language texts and spoken content.

6.1.1 SVL coverage. Figure 6 illustrates the coverage of the VocaBura method for each of the 12 vocabulary difficulty levels in the SVL corpus. We explored how many words in the SVL corpus are covered by the location-related word lists by randomly selecting L locations from the total of 3982 locations, where L was increased from 1 to 3982. The VocaBura method is able to cover 9949 out of the 12,000 words (82.9 percent) within the whole corpus when using all available location-related word lists (L=3982).

Fig. 6. The horizontal axis represents L and the vertical axis the number of words found in the SVL corpus for each difficulty level. As the number of words for each L changed depending on the combination of locations, the average value from 10,000 combinations is plotted with the exception of L=1 with 3982 combinations and L=3982 with one combination. The error bars represent the standard deviation.

Figure 7 illustrates the SVL coverage when setting the number of randomly selected locations to 100. This number roughly represents the number of locations with different location categories a user would pass when walking approximately 1500 meters within a large city (see: Chapter 6.2). The data shows that VocaBura can cover approximately 50 percent of the beginner level vocabulary within 100 locations, and maintains over 30 percent coverage up to SVL level 5. However, for advanced level L2 vocabulary (SVL 11 and 12), the coverage is less than 10 percent. For SVL level 1, 232 words were not included in any of the location-related word lists generated by Gensen Web (Figure 5). These words include English articles, pronouns, and basic verbs and adverbs.

6.1.2 New General Service List coverage. Figure 8 illustrates the coverage of the VocaBura method for the New General Service List corpora. We explored how many words in the corpora are covered by the location-related word lists by randomly selecting L locations from the total of 3982 locations, where L was increased from 1 to

¹⁶<http://www.newgeneralservicelist.org/>

Fig. 7. VocaBura coverage for the 12 difficulty levels in the SVL corpus with 100 locations.

3982. The VocaBura method is able to cover 6805 out of the 7360 words (92.5 percent) within the New General Service List corpora when using all available location-related word lists.

Figure 9 illustrates the New General Service List coverage when setting the number of randomly selected locations to 100. The data shows that VocaBura can cover approximately 70 percent of the beginner level Core General English Words (NGSL) vocabulary within 100 locations and nearly all of the Core Spoken (NGSL-S) list. For the specialized vocabulary lists, the coverage was approximately one third of each corpus (NAWL, TSL, BSL). Similarly to the SVL corpus, English articles, pronouns, and basic verbs and adverbs were not covered in the Core General English Words (NGSL) list.

Fig. 8. The horizontal axis represents L and the vertical axis the number of words found in each of the New General Service List corpora. As the number of words for each L changed depending on the combination of locations, the average value from 10,000 combinations is plotted with the exception of L=1 with 3982 combinations and L=3982 with one combination. The error bars represent the standard deviation.

Fig. 9. VocaBura coverage of New General Service List corpora with 100 locations.

6.2 Use Case of Long-term L2 Vocabulary Acquisition Potential with VocaBura

In this section, we present a use case of L2 vocabulary acquisition potential with VocaBura by using data from the real-life walking commute routes of two individuals working in a large city. For this analysis, we first identified the walking commute for the first and second author (hereafter Route A and Route B), and then extracted all location data from Google Maps along these routes. The routes did not overlap at any point and were unique in the sense that Route A was along a busier main street, whereas Route B followed a comparatively quieter side street in the same general area.

Route A measured at approximately 1050 meters one way and included a total of 122 unique locations available through Google Maps at the time of writing. These locations belonged to a total of 72 unique Google My Business categories. One location category, including three unique locations, was added to the My Business categories after February 2018, and we omitted these locations from the use case. The final data set for Route A included 119 unique locations belonging to 71 unique location categories.

The most common location categories were 'cafe' at 13 locations, 'convenience store' at 7 locations, and 'coffee shop' at 4 locations. The rest of the location categories included variations of the restaurant-category (e.g., 'Thai restaurant', 'Chinese restaurant') with one to two unique locations each, and various other location categories commonly found in a large city proper, such as 'ATM', 'auditorium' and 'post office'.

Route B measured at approximately 750 meters one way and included a total of 71 unique locations belonging to 50 Google My Business categories. One category representing one location was added after February 2018, and was omitted from this use case. The final data set for Route B included 70 unique locations belonging to 49 unique location categories. The most common location categories were 'parking lot' at 8 locations, and 'corporate office' and 'cafe' both at 4 locations. The rest of the location categories included various shops (e.g., 'home goods store', 'hair salon') and restaurants (e.g., 'steak house', 'Italian restaurant'). In total, there were 19 location categories that were shared between Route A and Route B. That is, Route A had 52 location categories not appearing on Route B, whereas Route B had 30 location categories not appearing on Route A.

Next, to investigate the potential L2 vocabulary coverage for Route A and Route B we used the process described in Figure 5 to create two route-specific word lists and calculated the vocabulary coverage for both routes in the SVL corpus and the New General Service List corpora. Figure 10 illustrates the number of unique and overlapping words for Route A and Route B for each SVL difficulty level. Figure 11 illustrates the number of unique and overlapping words on each route for the New General Service List corpora.

Fig. 10. Unique and overlapping coverage of English words on Route A and Route B in the SVL corpus. Coverage of 100 random locations as reference.

Fig. 11. Unique and overlapping coverage of English words on Route A and Route B in the New General Service List corpora. Coverage of 100 random locations as reference.

The data illustrated in Figure 10 reveals that the majority of overlapping words on the two routes are beginner level vocabulary (SVL1 - SVL5). However, both routes also include a considerable number of unique words to the route and the locations along it. Overall, this data suggests that VocaBura users could study approximately 400-500 unique English words up to SVL5-level on either of the routes defined in this use case. Similarly to the SVL corpus, overlapping words in the New General Service List corpora on these two routes are mainly in the NGSL corpus including the core English vocabulary. On average, Route A and Route B have unique coverage of 274 NGSL, 47 NGSL-S, 89 NAWL, 126 TSL and 173 BSL words in the New General Services List corpora (Figure 11).

When matched with the parameters of our Study 1, beginner level VocaBura users could theoretically cover the unique vocabulary in both the SVL corpus and the New General Service List corpora along either of the example routes within approximately one and a half months by studying 20 words on each working day (10 words per walking direction). Depending on the user's walking speed, either route would take approximately 10 to 15 minutes to complete one way.

7 DISCUSSION

In Study 1, Japanese beginner level English language learners used a context-aware audio-based L2 vocabulary learning method, VocaBura, to discover new English language words by accident while walking. Our findings indicated that they were able to retain significantly more new L2 words in their long-term memory compared to an audio-only baseline learning method with which they could fully focus on the learning task. Secondly, the participants reported as being aware of learning significantly fewer words with the proposed VocaBura method compared to the baseline audio-only learning method. In Study 2, we uncovered the level of L2 English vocabulary coverage achievable with the VocaBura method. Our results contribute to CTML and dual-coding theories, as well as increase our theoretical understanding on audio-based L2 incidental microlearning while walking. We further discuss the practical design implications arising from our two studies for the development of context-aware and audio-based L2 microlearning applications.

7.1 Contextual L2 Microlearning while Walking

Our results suggested that combining visual (i.e., appearance of a location) [3, 23, 24, 29, 37] and contextual cues [2, 8, 32] with audio-based verbal L1-L2 information improved the long-term retention of new L2 words for beginner level English language learners compared to audio-only L2 learning. Open-ended comments from our experiment participants also reflected how the interdependent visual and verbal information as well as the contextual cues from real locations helped them memorize the L1-L2 word pairs.

"It was easy to store information in your brain when you could associate the things that you saw with your eyes to things that you heard." (participant ID12)

"I thought it was easy to remember the words since they were related to real locations." (participant ID3)

Firstly, this result supports the assumptions of CTML and dual-coding theories [23, 24], suggesting that the combination of visual and verbal information improves the retention of L2 words [37]. However, our present results do not answer whether or not these added cues aided the L2 learners in the immediate recall test (Figure 4). Secondly, our results contribute to the literature on the efficacy of contextual microlearning methods for L2 vocabulary acquisition. Previous works have reported on the number of learned words [8] as well as the recall of new L2 vocabulary items [4] studied with a text-based contextual microlearning application. However, to best of our knowledge, our study is the first to report on both immediate and delayed vocabulary recall of L2 words learned with an audio-based contextual microlearning method while the learners are walking.

7.2 L2 Microlearning by Accident

According to Krashen's Monitor Model, incidental L2 vocabulary learning is subconscious [19–21]. That is, L2 learners may learn new words by accident when watching a TV show or overhearing a conversation while having no awareness of acquiring the new vocabulary [31]. In this work, our goal was to explore to what extent L2 vocabulary learning with our proposed VocaBura learning method would occur by accident while learners are focused on another task - walking along a route.

Our results suggested that beginner level English language learners have significantly lower awareness of vocabulary acquisition having occurred when using our proposed method compared to a baseline audio-only learning method. Approximately 40 percent of the L2 vocabulary that the participants recalled correctly in the immediate test were words that they learned without awareness while focusing on walking and using the VocaBura learning method. Conversely, the participants reported as having no awareness of learning for only 10 percent of the words they studied with the baseline method and recalled correctly in the immediate test. This indicated that while the learners were focused on walking, they were unaware of learning L2 words related to the locations they passed [16, 17, 31], but were regardless able to recall these words without awareness of learning having occurred.

These results can be further explained by open-ended comments from the participants regarding the VocaBura learning task.

"I was more interested in the scenery than the English words." (participant ID13)

"There was a lot of information to process when walking past a new location, so it was quite difficult to focus on the words." (participant ID14)

The walking route used during the experiment was unfamiliar to the majority of the participants and this may have affected their attention during the VocaBura learning task. On one hand, some participants were focused on the new scenery, as would real users who are familiarizing themselves with their surroundings. On the other hand, the participants also reported being focused on processing the rich information from the unfamiliar real-life locations rather than focusing on the words they heard while walking. Thus, the participants' self-reports further suggest that they were focused on the task of walking and scanning their surroundings rather than the L2 vocabulary learning task. However, further evidence is needed to explore how users' familiarity with their walking route might affect their attention and the recall of L2 words.

These results suggest that the VocaBura learning method may at least partially promote L2 learning that is akin to incidental learning. Previous works suggest that readers recall more words they learned incidentally rather than intentionally [1, 15], and that audio-based learning may be beneficial for incidental vocabulary acquisition [22, 35]. However, we are hesitant to make definite assumptions regarding the mental processes related to L2 learning with the VocaBura method without further experimentation. Hence, we conclude that while the VocaBura method requires learners to be focused on an unrelated task during L2 vocabulary study, they are still able to acquire new L2 vocabulary by accident and retain this vocabulary in long-term memory better than with a traditional audio learning method.

7.3 Potential L2 Vocabulary Coverage of Automated VocaBura Learning Method

In Study 2, our aim was to explore to what extent an automated VocaBura learning method would cover English vocabulary for L2 learners. That is, how many words are L2 learners able to study with our proposed method, and what difficulty level are these words.

While our proposed method is capable of reaching a high level of coverage in the two L2 English corpora (ALC SVL and New General Service List) used for our evaluation, reaching the high coverage level would require learners to walk past thousands of different locations. In our second analysis, we set the number of locations to 100 which roughly corresponds to the number of possible locations within a walking distance in a large city (see: Chapter 6.2). The results from this analysis suggested that by using our proposed method beginner and intermediate level L2 English learners could study between 30 to 50 percent of the required vocabulary to reach the next level of proficiency in the ALC SVL. However, the VocaBura method offers less coverage for advanced learners since as the difficulty of L2 words increases their usage in English language Wikipedia articles related to locations and location categories decreases. To lesser extent, this is also the case for specialized academic and business vocabulary, where our proposed method can cover approximately one third of the New General Service List specialized corpora. However, the VocaBura method is able to cover approximately 70 percent of the NGSL Core General English words and nearly all of NGSL-S Core Spoken words, which combined correspond to approximately 87 to 94 percent of basic English language vocabulary used in texts and spoken dialogues¹⁷.

In our third analysis, we introduced two real-life walking commutes as use cases. The data extracted from these routes followed a similar pattern as with our theoretical 100 location models, where the majority of words covered by our method were of beginner and intermediate level (Figure 10). While there was some overlap in location categories between the two example routes, the majority of locations were unique to each route. Subsequently, our data showed that while there was overlap in terms of L2 word coverage along the example routes, they also

¹⁷<http://www.newgeneralservicelist.org/>

included considerable number of vocabulary that were unique to the route for both the SVL corpus and the New General Service List corpora (Figure 10, Figure 11). Our data showed that learners using the VocaBura method could potentially study approximately 400-500 unique L2 words up to SVL5-level and approximately 700 unique words in the New General Services list corpora in one and a half months by walking either of the commute routes on every working day. Moreover, this analysis revealed that beginner and intermediate level L2 learners could increase the number of new vocabulary they are able to study with VocaBura by simply modifying their daily walking commute. This real-life use case further highlights the potential of our proposed learning method in a large city, but future studies should also analyze the vocabulary coverage in smaller cities and less populated areas.

Altogether, these results demonstrate the efficacy of our approach for beginner and intermediate level English language learners. We argue that the VocaBura learning method can be useful for both acquiring new L2 vocabulary while walking, but also for reviewing vocabulary already studied in a classroom setting. We believe that using VocaBura alongside formal language study could increase the speed in which beginner and intermediate level English learners reach advanced language proficiency, and ultimately fluency in their second language.

7.4 Design Implications

Our results indicated a number of opportunities for design to improve the efficacy of contextual L2 microlearning applications.

7.4.1 L2 vocabulary learning as a daily routine. The first design implication arising from our results is the opportunity to implement audio-based and context-aware L2 vocabulary learning in the background of users' daily routines. That is, L2 learners who walk the same route to work or school every day could make use of the familiar locations and contextual audio-based microlearning to study and review new L2 vocabulary. For example, the VocaBura method could be integrated with a music player so that the music would be gradually muted as the learner approaches a location, as opposed to the audible cue used in our Study 1 experiment. This type of application could be used in addition to classroom work and textbook study, where beginner and intermediate level learners could increase the speed of acquiring new L2 vocabulary by discovering new words not yet covered in formal lessons, or by reviewing already covered vocabulary in the context of use.

Secondly, the system we introduced in this work did not yet include a functionality similar to spaced repetition of flashcards used in MicroMandarin or Anki applications [9, 13]. As is evident from our immediate recall results, beginner level learners may not be able to pick up all L2 words during one session with VocaBura, but would rather have to revisit already heard L2 vocabulary items on a later date. Conversely, already learned words should be repeated in longer intervals more as 'refreshers' (i.e., spaced repetition). One way to achieve this functionality would be to design and implement a vocabulary test application that the learners could use on their mobile phone after finishing their commute or walk. Results from this immediate vocabulary test could be then used to inform a spaced repetition algorithm on which words to omit as learned and which ones to include as not learned the next time the learner walks past the same location or location category.

Thirdly, our results suggested that the participants learned at least some of the L2 vocabulary without awareness, or by accident, when using VocaBura. While our experimental results indicated that L2 learners were able to retain more vocabulary items in their long-term memory with our proposed approach, lack of awareness about vocabulary acquisition may also introduce detrimental effects to L2 learners' motivation for prolonged language study. That is, if a learner's perception is that they did not acquire new vocabulary during their walk or commute while using a L2 microlearning application similar to VocaBura (i.e., no awareness of learning), they may cease to use the application prematurely. Hence, it is important to design for a feedback method which ensures users that they are making progress and acquiring new L2 vocabulary. Similarly to spaced repetition of L2 words, this

functionality could be achieved by implementing a vocabulary recall test which then provides feedback on the results and overall learning progress.

7.4.2 Design for complementing motivations. Second language learning requires time investment and motivation from the L2 learner to keep up on daily basis [5, 8, 16]. Daily exercise is a similar activity in the sense that people need to adhere to a steady routine in order to keep up with their health or exercise goals. Instead of using separate apps for language learning and exercise support, our proposed approach offers the opportunity to combine the motivations for long-term engagement in both activities in one application.

Since our results support the notion that L2 learners can acquire contextually relevant vocabulary while simultaneously focusing on physical exercise such as walking, we recommend designing contextual L2 microlearning applications to complement daily exercise (e.g., running, walking) or sightseeing. Joggers and runners who might find it difficult to meet their exercise goals could find new motivation to jog if they could meet their L2 study goals simultaneously and track both their L2 study and exercise progress in one application. One way to achieve this type of application could be to combine the VocaBura system with an exercise application for either a smartwatch or a smartphone that uses the GPS functionality to track the joggers' progress.

Similarly, foreign students moving to a new country could benefit from an application that simultaneously supports sightseeing or getting familiar with the new surroundings (e.g., location-based audio guidebook) and L2 vocabulary learning during these activities or while transiting from one location to another (i.e., dead time). Combining language learning and sightseeing in one application could motivate foreign students to further their L2 vocabulary study since they would be simultaneously able to discover new and exciting surroundings in their new host country and locale.

7.4.3 Improving spatial contiguity in contextual microlearning with augmented reality (AR). One design implication arising from Study 1 is the need to support spatial contiguity in contextual microlearning applications. We recommend future development to consider secondary cues to either indicate the direction of a contextually relevant location, or artificially shorten the distance between the learner and the location. One way to achieve this would be to *blow up* a location image as an overlay to user's AR glasses so as to achieve an illusion that the location is nearby when studying a related L2 word.

7.4.4 Increasing location-related L2 vocabulary corpus coverage by word stringing. The results from our second study showed how an automated method for creating location-related L2 corpora omitted some beginner level and advanced level vocabulary items. On one hand, beginner level words were omitted due to their general use independent from context, while advanced level vocabulary was not found in location-related Wikipedia articles.

Our suggestion for increasing the coverage of automatically generated location-related L2 corpora is to automatically string together already learned words with either beginner or advanced level words that are not directly related to any location. One way to achieve this without breaking the contextual relevance of a word pair would be to automatically string already learned nouns within the corpus with an advanced level adjective. For example, learners may have studied the word waitress (SVL level-2 word) related to a restaurant location. The system could then string the word waitress with the word impudent (SVL level-12 word), and thus automatically increase the coverage of the location-related corpus. Similarly, beginner level words not initially covered in the corpus could be stringed with already learned location-related words to form sentences as the L2 learners advance in their studies.

7.5 Limitations and Future Work

In future work, we are interested in improving and expanding the VocaBura learning method to more naturalistic contexts. Firstly, we are planning on fully automating the L1-L2 word pair generation so that learners may make use of the application independent of their native language, L2 they are studying, or their locale. Our

plan is also to design a version of the VocaBura application for exercise support and evaluate its efficacy in a longitudinal study. In this future work, we are interested in investigating aspects of learning motivation and long-term commitment to L2 vocabulary learning with our proposed approach independently, as well as in combination with traditional classroom study. Furthermore, while the learners in our experiment studied each word only once, we are interested in exploring ways to optimize a spaced repetition algorithm for the VocaBura learning method.

Secondly, our system design introduced in Study 1 used a 25-meter radius around a location as a physical distance threshold to play the L1-L2 audio word pairs. This design choice was initially made based on preliminary experiment results to compensate for inaccuracies in detecting learners' GPS location while they were mobile [14], but may have still affected the spatial contiguity during the VocaBura learning task. In future work, we are interested in exploring ways, such as the use of mobile AR, to improve spatial contiguity with visual, contextual and audio learning cues while L2 learners are walking.

Thirdly, one limitation in the design of Study 1 was the order of participants' exposure to VocaBura and baseline learning methods. Due to the absence of counterbalancing between the conditions, the participants' performance in the *baseline condition* may have been higher than expected for novice users. This is due to the fact that they were able to acclimate themselves to the artificial voice used for narrating the L1-L2 words in the VocaBura condition, and become familiar with the vocabulary learning and testing process in general. Open-ended comments from participants also reflected this assertion, where some participants felt they could not understand the L1 (i.e., Japanese) artificial voice.

"There were places where I could not understand what the Japanese [voice] was saying while walking." (participant ID8)

"I thought it might be great to memorize words while walking, but many times I could not catch what the Japanese [voice] was saying." (participant ID5)

However, it is important to note that this limitation in the experimental design actually worked against our hypotheses, and future works with fully counterbalanced experimental conditions may result in lower immediate and delayed recall test scores with the *audio-only baseline learning method* than reported in this study.

Lastly, as mentioned in Section 6 as well, analyses in Study 2 did not consider the level of contextual relatedness of words to a specific location. That is, while the current VocaBura method extracts word lists based on the location information available in Wikipedia, it does not directly consider how well individual words within these lists relate to a specific shop, restaurant, park or other location. In future work, we are developing a method to link L2 words to specific locations by not only considering the business category but the individual business or location as well.

8 CONCLUSION

In this paper we introduced VocaBura, a context-aware audio-based vocabulary learning system which enables second language vocabulary learning during users' dead time, such as while walking to and back from work or school. We conducted a controlled vocabulary learning experiment with Japanese beginner level English language learners to evaluate the efficacy of our proposed approach. Our experimental results suggested that beginner level English learners were able to retain significantly more English language words with the proposed method compared to a baseline audio-only learning method. Furthermore, we showed that since learners had to focus on walking while studying with the VocaBura method, they acquired a significant portion of the new second language vocabulary without awareness, or by accident. Results from our second study detailed to what extent a fully automated VocaBura system can support beginner and intermediate level English language learners' vocabulary acquisition and complement formal classroom study. Our findings contribute to second language

acquisition and microlearning literature, and suggest several ways that future language learning tools could better support contextual second language microlearning.

ACKNOWLEDGMENTS

This research was supported by JST PRESTO Grant Number JPMJPR1658.

REFERENCES

- [1] Jameel Ahmad. 2012. Intentional vs. incidental vocabulary learning. *ELT Research Journal* 1, 1 (2012), 71–79.
- [2] John Seely Brown, Allan Collins, and Paul Duguid. 1989. Situated Cognition and the Culture of Learning. *Educational Researcher* 18, 1 (1989), 32–42. <https://doi.org/10.3102/0013189X018001032>
- [3] Dorothy M. Chun and Jan L. Plass. 1996. Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal* 80, 2 (1996), 183–198. <http://www.jstor.org/stable/328635>
- [4] David Dearman and Khai Truong. 2012. Evaluating the implicit acquisition of second language vocabulary using a live wallpaper. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 1391–1400. <https://doi.org/10.1145/2207676.2208598>
- [5] Zoltan Dornyei and Istvan Otto. 1998. Motivation in action: A process model of L2 motivation. In *Working Papers in Applied Linguistics*, Vol. 4. Thames Valley University, 43–69. <http://eprints.nottingham.ac.uk/39/>
- [6] Darren Edge, Kai-Yin Cheng, Michael Whitney, Yao Qian, Zhijie Yan, and Frank Soong. 2012. Tip Tap Tones: Mobile microtraining of Mandarin sounds. In *Proceedings of the 14th International Conference on Human-computer Interaction with Mobile Devices and Services (MobileHCI '12)*. ACM, New York, NY, USA, 427–430. <https://doi.org/10.1145/2371574.2371640>
- [7] Darren Edge, Stephen Fitchett, Michael Whitney, and James Landay. 2012. MemReflex: Adaptive flashcards for mobile microlearning. In *Proceedings of the 14th International Conference on Human-computer Interaction with Mobile Devices and Services (MobileHCI '12)*. ACM, New York, NY, USA, 431–440. <https://doi.org/10.1145/2371574.2371641>
- [8] 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 (CHI '11)*. ACM, New York, NY, USA, 3169–3178. <https://doi.org/10.1145/1978942.1979413>
- [9] Damien Elmes. 2008. Anki: powerful, intelligent flashcards. (2008). <https://apps.ankiweb.net/>
- [10] Shogo Fukushima, Ari Hautasaari, and Takeo Hamada. 2019. Second language vocabulary learning while walking. In *Proceedings of the 10th Augmented Human International Conference 2019 (AH2019)*. ACM, New York, NY, USA, Article 44, 2 pages. <https://doi.org/10.1145/3311823.3311866>
- [11] Thomas J. Garza. 1991. Evaluating the use of captioned video materials in advanced foreign language learning. *Foreign Language Annals* 24, 3 (1991), 239–258. <https://doi.org/10.1111/j.1944-9720.1991.tb00469.x>
- [12] Gerhard Gassler, Theo Hug, and Christian Glahn. 2004. Integrated micro learning—An outline of the basic method and first results. *Interactive Computer Aided Learning* 4 (2004), 1–7.
- [13] Robert Godwin-Jones. 2010. Emerging technologies from memory palaces to spacing algorithms: approaches to second language vocabulary learning. *Language, Learning & Technology* 14, 2 (2010), 4.
- [14] Takeo Hamada, Shogo Fukushima, and Ari Hautasaari. 2018. Incidental second language vocabulary learning method with location-based audio. *IEICE Technical Report (In Japanese)* MVE2017-71 (2018), 29–33.
- [15] Jan H. Hulstijn. 1992. *Retention of Inferred and Given Word Meanings: Experiments in Incidental Vocabulary Learning*. Palgrave Macmillan UK, London, 113–125. https://doi.org/10.1007/978-1-349-12396-4_11
- [16] Jan H. Hulstijn. 2001. *Intentional and Incidental Second Language Vocabulary Learning: A Reappraisal of Elaboration, Rehearsal and Automaticity*. Cambridge University Press, 258–286. <https://doi.org/10.1017/CBO9781139524780.011>
- [17] Jan H. Hulstijn. 2008. *Incidental and Intentional Learning*. John Wiley & Sons, Ltd, Chapter 12, 349–381. <https://doi.org/10.1002/9780470756492.ch12>
- [18] Daesang Kim and David A. Gilman. 2008. Effects of text, audio, and graphic aids in multimedia instruction for vocabulary learning. *Educational Technology and Society* 11 (07 2008), 114–126.
- [19] Stephen Krashen. 1982. *Principles and Practice in Second Language Acquisition*. Oxford Pergamon.
- [20] Ronald Leow and Luis Cerezo. 2016. Deconstructing the I and SLA in ISLA: One curricular approach. *Studies in Second Language Learning and Teaching* 6, 1 (2016), 43–63. <https://doi.org/10.14746/ssllt.2016.6.1.3>
- [21] Ronald P. Leow and Celia C. Zamora. 2017. *Intentional and Incidental L2 Learning*. Routledge, Chapter 3. <https://doi.org/10.4324/9781315676968.ch3>
- [22] Jonathan Malone. 2018. Incidental vocabulary learning in SLA: Effects of frequency, aural enhancement, and working memory. *Studies in Second Language Acquisition* 40, 3 (2018), 651–675. <https://doi.org/10.1017/S0272263117000341>

- [23] Richard E. Mayer. 2009. *Multimedia Learning* (2 ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511811678>
- [24] Richard E. Mayer and Valerie K. Sims. 1994. For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology* 86, 3 (1994), 389–401. <https://www.learntechlib.org/p/79160>
- [25] Hiroshi Nakagawa and Tatsunori Mori. 2003. Automatic term recognition based on statistics of compound nouns and their components. *Terminology* 9, 2 (2003), 201 – 219. <http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=12580627&lang=ja&site=ehost-live>
- [26] Tatsuya Nakata. 2015. Effects of feedback timing on second language vocabulary learning: Does delaying feedback increase learning? *Language Teaching Research* 19 (07 2015), 416–434. <https://doi.org/10.1177/1362168814541721>
- [27] Allan Paivio. 1986. *Mental Representations: A Dual Coding Approach*. Oxford University Press.
- [28] Allan Paivio and Alain Desrochers. 1980. A dual-coding approach to bilingual memory. *Canadian Journal of Psychology/Revue canadienne de psychologie* 34 (1980), 388–399. Issue 4. <https://doi.org/10.1037/h0081101>
- [29] Jan L. Plass, Dorothy M. Chun, Richard E. Mayer, and Detlev Leutner. 1998. Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *Journal of Educational Psychology* 90, 1 (1998), 25–36. <https://www.learntechlib.org/p/87147>
- [30] Falcon Dario Restrepo Ramos. 2015. Incidental vocabulary learning in second language acquisition: A literature review. *Profile: Issues in Teachers' Professional Development* 17, 1 (2015), 157–166. <https://doi.org/10.15446/profile.v17n1.43957>
- [31] Angelika Rieder. 2003. Implicit and explicit learning in incidental vocabulary acquisition. *Views* 12 (01 2003).
- [32] Jenny R. Saffran, Elissa L. Newport, Richard N. Aslin, Rachel A. Tunick, and Sandra Barrueco. 1997. Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science* 8, 2 (1997), 101–105. <https://doi.org/10.1111/j.1467-9280.1997.tb00690.x>
- [33] Endel Tulving and Donald M. Thomson. 1973. Encoding specificity and retrieval processes in episodic memory. *Psychological Review* 80, 5 (1973), 352–373. <https://doi.org/10.1037/h0020071>
- [34] Christian David Vazquez, Afika Ayanda Nyati, Alexander Luh, Megan Fu, Takako Aikawa, and Pattie Maes. 2017. Serendipitous language learning in mixed reality. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*. ACM, New York, NY, USA, 2172–2179. <https://doi.org/10.1145/3027063.3053098>
- [35] Karina Vidal. 2011. A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning* 61, 1 (2011), 219–258. <https://doi.org/10.1111/j.1467-9922.2010.00593.x>
- [36] Stuart Webb. 2007. The effects of repetition on vocabulary knowledge. *Applied Linguistics* 28 (03 2007). <https://doi.org/10.1093/applin/aml048>
- [37] Makoto Yoshii. 2006. L1 and L2 glosses: Their effects on incidental vocabulary learning. *Language Learning and Technology* 10 (09 2006).