Fukushima *Research and Practice in Technology Enhanced Learning* (2019) 14:16
https://doi.org/10.1186/s41039-019-0109-0

Research and Practice in Technology Enhanced Learning

RESEARCH Open Access

EmoTan: enhanced flashcards for second language vocabulary learning with emotional binaural narration

Shogo Fukushima[1,2]

Correspondence:
shogo@nae-lab.org
[1]Interfaculty Initiative in Information Studies, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku 1138656, Tokyo, Japan
[2]PREST, Japan Science and Technology Agency, 4-1-8 Honcho, Kawaguchi, 3320012 Saitama, Japan

Abstract

We report on the design and evaluation of a flashcard application, enhanced with emotional binaural narration to support second language (L2) vocabulary learning. Typically, the voice narration used in English vocabulary learning is recorded by native speakers with a standard accent to ensure accurate pronunciation and clarity. However, a clear but monotonous narration may not always aid learners in retaining new vocabulary items in their semantic memory. As such, enhancing textual flashcards with emotional narration in the learner's native language can foster the retention of new L2 words in episodic rather than semantic memory as greater emotive expression reinforces episodic memory retention. We evaluated the effects of binaural emotive narration with traditional textual flashcards on L2 word retention (immediate and delayed) in laboratory experiments with native Japanese-speaking English learners. Our results suggest that the learners were able to retain approximately 60% more L2 words long-term with the proposed approach compared to traditional flashcards.

Keywords: Emotion, Narration, Binaural recording, Flashcard, Computer-assisted language learning, Vocabulary learning

Introduction

Vocabulary is fundamental to English language learning as its lack impedes language-related activities, such as conversing and multi-media content comprehension (Nation 2006; van Zeeland and Schmitt 2013). To increase vocabulary knowledge, a learner has to encounter the language repeatedly in daily micro-times, and hence, mobile language learning devices are becoming important. The number of mobile language learning studies has been increasing annually (Hwang and Fu 2019), and it has been proved that these applications can enable the learner to improve their language skills (Hwang and Wu 2014). To learn new English words with these mobile devices, learners generally memorize words while listening to the pronunciation in online dictionaries (Weblio 2005; Goo dictionary 1999; Jayme Adelson-Goldstein 2015; American Heritage Dictionary. Online dictionary 1969) or flashcard applications (mikan 2014; Smart Language Apps Limited 2015). The pronunciations recorded by native English speakers have no specific accent and, therefore, are accurate and easy to comprehend. Furthermore, although the vocalization is adequate for a learner's verification of his/her pronunciation of already-known words, it is not sufficient for understanding and memorizing new words (Wright et al. 2013).

To redesign vocalization for introducing new words, some commercial off-the-shelf language learning applications have been released. Moetan is a pioneering work (Moetan 2008) that alters the vocalization to the voice of a game character. Released for smartphones and gaming devices, this application allows learners to learn a word through sample sentences and voice. In addition, other stock voice applications that continuously use professional voice actors have also been released (Hiroshi 2014; Ogura H. 2014). It is suggested that these game-based type language applications are useful for maintaining learning motivation or engagement (Hung et al. 2018). However, these applications use monaural voice and are an attempt at simply alternating the voice actor vocalizing the flashcard. Moreover, it is not enough for understanding the meaning of the target words or retaining the meaning of the word in long-term memory.

Research in cognitive and affective neuroscience has suggested that emotional arousal enhances the long-term retention of human memory (Kensinger and Corkin 2003; Kleinsmith and Kaplan 1964; McGaugh 2003; Phelps et al. 1997). Accordingly, in this paper, we propose a method—"EmoTan"—to incorporate emotional stimulation into English vocabulary learning in flashcard applications on mobile devices. In Japanese, the word "Emo" means emotion and "Tan" means vocabulary. When an emotional stimulus is presented to a person, the experience is etched in the memory. Therefore, it is desirable that the contents of emotional stimuli reflect both the basic and fine meanings of words. To realize this, our method employs narration. Specifically, a story of several seconds in length is narrated to the learner, who can thereby perceive the contextual and refined meanings of the words. To make it easier to induce emotional arousal in the story, we increased the sense of immersion in the story with binaural recording technology. Moreover, we composed the story from the first person's perspective. The three-dimensional coordinates of the sound and movement could also be used as additional memory trace. Consequently, the learner can memorize the words in his/her emotionally rich episodic memory rather than semantic memory. For example, for the term "aerate," which means "to introduce air into (something)," a narration is recorded while blowing into a dummy head microphone (Fig. 1 (left, center)). Therefore, the learner can more deeply comprehend the meaning of the word when accompanied by the blowing sound and emotion-triggered bodily changes (Fig. 1 (right)).

To store the knowledge of vocabulary in long-term memory, word repetition and multiple study opportunities are necessary (Cull 2000). In non-English-speaking countries,

Fig. 1 Left: A voice actor is blowing into a dummy head microphone to express the meaning of the word "aerate," which means "to introduce air into (something)." Center: For the word "rumple," which means to mess up, for example, someone's hair, a wig was attached to the dummy head microphone for representing a realistic experience. Right: A learner experiencing the meaning of "aerate" through the recorded voice using a general earphone and a tablet

it is especially important that opportunities are provided for learners to practice or otherwise encounter the words outside English classes. Furthermore, a working adult has minimal time to spend on language learning. Therefore, the use of narration, which promotes word memorization, is socially important as well. In short, the key contributions of the proposed approach are as follows:

- We designed narrations to memorize English words in the emotionally rich episodic memory and not semantic memory, and we described the method of producing the narration.
- The effectiveness of the narration using flashcard application was evaluated by comparing our method to the traditional pronunciations of English words. The result implies that learning by the proposed narration method makes it significantly easier to remember an English word and its corresponding translation. However, note that it is not yet clear which variable of this narration affects this result in the present study.
- We show that our proposed method could aid in retaining more English words into long-term memory than traditional pronunciation by applying our method to a general English learning style, in which the learners freely memorize words using a typical flashcard application on a tablet until they are satisfied.

Related work

Memory-enhancing effect of emotion

Research in cognitive and affective neuroscience has shown that emotive expression enhances the long-term retention of human memory (McGaugh 2003). For instance, Kensinger and Corkin demonstrated that emotive words (e.g., funny, victim, error) are more readily recalled than non-emotive words (e.g., switch, locate, habit) (Kensinger and Corkin 2003). For a paired-associate learning format, Kleinsmith and Kaplan showed that nonsense-syllable-paired associations learned under a high arousal state produced nearly permanent memorization (Kleinsmith and Kaplan 1964). Furthermore, the long-term memorization of a word is likely to be retained merely by placing a neutral word that does not have an emotional meaning into an emotional context (Phelps et al. 1997).

An emotional state can be identified under a circumplex model of affect (emotion), with the horizontal axis representing the valence dimension and the vertical axis representing the arousal dimension (Russell 1980). The arousal dimension is regarded as key to the memory-enhancing emotion effect. Hamann et al. (1999) showed images to subjects to evoke various emotions. High-arousal images were memorized, while valence ones were not. The researchers conducted experiments, showing images to subjects to evoke various emotions, and demonstrated that arousal pictures were better remembered than valance (pleasant or unpleasant) ones. They also demonstrated that pictures with a high galvanic skin response, indicative of high arousal level, were better memorized than others (Hamann et al. 1999).

The hippocampus and peripheral limbic system are believed to be involved in the promotion of long-term memory by means of emotions (Hamann et al. 1999; LaBar and Cabeza 2006). Particularly, the secretion of noradrenaline in the amygdala acts on the most recently engaged synapse of the brain and is involved in the consolidation of all types of learning information (Hamann et al. 1999; LaBar and Cabeza 2006). Noradrenaline in the amygdala is secreted by stimulation due to excitement and stress. Therefore, for

example, if a memory test is performed during mild exercise, such as grasping a force meter, the arousal level will increase and the memory may be reinforced (Coles and Tomporowski 2008).

To apply the memory retention effect of emotion to vocabulary learning, it is better to use this retention effect not only for emotive words, such as in the study of Kensinger and Corkin (2003), but also for non-emotive words. For that, it is necessary to stimulate human emotions using external emotional stimulation while learning non-emotive words. International Affective Digitized Sounds (IADS) (Bradley and Lang 2007) and the International Affective Picture System (IAPS) (Lang et al. 2008) have been widely used as emotional stimuli in cognitive and affective neuroscience studies. Nevertheless, an insufficient number of sounds or sound effects exist compared to the size of the English vocabulary, making it difficult to find enough emotional stimuli that match the meaning of English words. Furthermore, since it is a pure experimental stimulus, learning becomes experimental, so it is not practical in assuming a case of daily vocabulary learning.

The major technique to evoke emotions with voice and sound are 3D sound attraction and Autonomous Sensory Meridian Response (ASMR). These techniques create emotive specific situations or stories by using binaural recording technology. Binaural recording is a method of creating a 3D stereo sound sensation for the listener which is similar to actually being in the room with the performers. This effect is created using a dummy head mannequin that is outfitted with a microphone in each ear. This has been applied not only for amusement parks (3D sound attraction of Joypolis 2016), but also for controlling moods and chronic pain (Barratt and Davis 2015). However, there has been no methodology introduced yet for applying this binaural emotional technology to English vocabulary learning.

Experience-based vocabulary learning

To acquire vocabulary in long-term memory, it is important to deeply understand the meaning of a word and repetitively learn the word to memorize it. To promote word understanding, there is a system that uses pictures (Jayme Adelson-Goldstein 2015) and sentences (Suzuki 2000). However, to promote a more efficient and effective use of vocabulary learning, learning within a brief time spaced throughout the day using mobile devices (micro-learning) has been studied (Cavus and Ibrahim 2009; Gassler et al. 2004). Additionally, there are mobile applications that guide the appropriate review timing based on the Ebbinghaus forgetting curve (Nakata 2015; Luis and von Severin 2011). Nonetheless, executing several repetitions remains necessary.

To mitigate mechanical repetition in vocabulary learning, learning as an experience or episodic memorization has also been actively studied (Nessel and Dixon 2008; WEARABLE LANGUAGE TEACHER ELI 2017). Episodic memory is the memory of an individual's experiences. It is indexed in a time-space context, such as a location and the surrounding environment. On the other hand, semantic memories are defined as memorized knowledge, where the temporal and spatial information in the learning no longer exist. English Learning Intelligence (ELI) (WEARABLE LANGUAGE TEACHER ELI 2017) is a learning method that involves compiling a diary in English, that is, a written record of daily activities and conversations, which involves the construction of English sentences. Additionally, a vocabulary learning application that uses context information for memorization was developed (Al-Mekhlafi et al. 2009; Dearman and Truong 2012;

Hsieh et al. 2007; Ogata and Yano 2004). However, the vocabulary is heavily dependent on the context. To cope with this limitation, the use of multimedia content, such as movies, videos, and television, has been studied. ViVo, for example, is a dictionary that extracts short video clips from movies and television based on keywords from subtitles and combines them with images for learning (Zhu et al. 2017).

It is thus becoming possible to search for the usage of a target word that the learner wants to remember from daily life and multimedia contents as example sentences (WEARABLE LANGUAGE TEACHER ELI 2017; Al-Mekhlafi et al. 2009; Dearman and Truong 2012; Hsieh et al. 2007; Ogata and Yano 2004) or videos (Zhu et al. 2017). Although these sentences include the target word, they were not designed to represent the meaning of the target word. Therefore, it is unclear whether they are sufficient for understanding and memorizing new words. Furthermore, the audio displays for presenting sentences in previous works were seldom in stereophonic sound using the binaural recording technique and were not designed to evoke emotions in learners from the experience. In this paper, we describe a method of producing the audio contents that represent the meaning of the English words and evoke emotion at the same time.

Narration design

To incorporate emotional stimulation into English vocabulary learning, we focus on the pronunciation used in online dictionaries and flashcard applications on mobile phones. While it may be possible to use international standard emotional sounds, such as IADS (Bradley and Lang 2007), there are limitations on the types of arousal effect sounds. It is thus difficult to respond to the various types of words. For that reason, we originally created the narration of the arousal emotion corresponding to the word.

When an emotional stimulus is presented to a person, the experience episode is strongly memorized. Therefore, it is desirable that the contents of emotional stimuli reflect both the basic and fine meanings of words. Additionally, if one episode is close to another episode in time, memory may interfere (Loftus 1996). Therefore, we attempted to create different scenes and stories for each word. Furthermore, the content needed to represent an emotional arousal stimulus.

To make it easier to induce emotions, we increased the sense of immersion in the story with binaural recording technology. Moreover, we composed the story from the first-person perspective. The three-dimensional coordinates of the sound and movement could also be used as additional memory trace. The production steps were as follows:

- Selection of narrator;
- Selection of words;
- Creation of expressions and scripts;
- Audio recording.

To create impressive emotional narrations, expressive voice actors were adopted as narrators under the supervision of English native speakers (supervisors) with English teaching experience. To select a narrator, we first conducted an audition based on three criteria: *accuracy of pronunciation*, *ease of being heard*, and *expressive voice*. The narration used for the audition was a sample submitted by each voice acting office; accordingly, each voice actor spoke various lines in English. Specifically, an experiment was conducted by the supervisor as follows. We asked five female and three male candidates from the voice

acting offices of four companies to participate in the audition. The supervisor listened to the narration sample and answered a questionnaire. Based on the responses, we chose the voice actor with the highest average score.

First, we selected 1000 words other than the most frequent 9000 word families of the British National Corpus to eliminate from the experiment the influence of prior knowledge. Next, we selected only words of four, five, and six letters (approximately 370 words) since experiments in related works were performed using these word lengths (Nakata 2015; Webb 2007). Additionally, considering the compatibility with our method, we selected words having an action or sound meaning, which are known to influence the user psychology (approximately 100 words). According to their meanings, 100 words were grouped into ten categories: person, emotion, contact, communication, food, object, social, environment, impression, and sound effect. These categories were defined by referencing the 25 nouns and 15 verbs at the top of the Princeton WordNet hierarchy. We selected approximately three words from each category (30 wordsin total).

To maximize the memory trace in episodic memory, we established specific situation parameters of "when," "where," "who," and "how." However, it seemed difficult to sufficiently express the nuance of the meanings of words only by emotionalizing the pronunciation. We therefore added several seconds of a short story that expressed the meaning of the given word in its context.

Three types of arousal emotions were considered: pleasant arousal expression, unpleasant arousal expression, and neutral arousal expression. Based on these emotions, we brainstormed the representation of stories in collaboration with a content production company (InfoBurn Co., Ltd.). The categorized expressions were obtained and grouped as follows: (1) one person, expressions in which a voice actor acts alone; (2) two persons, expressions of actions in with two or more voice actors; (3) approach, expressions in which the voice actor approaches the listener; (4) touch, expressions in which the voice actor touches the listener; (5) circulate, expressions in which the voice actor circulates around the listener; and (6) together, expressions experienced with voice actors. Note that in case of (2) two persons, the voice actor plays two roles. As she is a professional voice actor, she can perform as characters with different voice tones. The expressions were selected to ensure a balance between these types. The script was determined based on the above points.

We recorded the narration in a binaural recording studio (Fig. 2). Upon recording, we subjectively judged whether each expression was reproduced or whether emotional arousal was evoked. Regarding emotion, we measured the skin conductance response (SCR), which is the electrical conductance of the palm that varies according to moisture level. This moisture change is controlled by the sympathetic nervous system. Thus, SCR is used as an index of emotional arousal (Vetrugno et al. 2003). If a SCR change was not registered, we revised the story and recorded it again. The SCR measuring device (DA-3, VEGA Systems Inc.) was used based on the recommendation of the Society for Psychophysiological Research, and appropriate preprocessing was applied.

The word list is shown in Table 1. The first column represents an expression category, and the second represents words and the length of narration. In the case of the word "swig" (expression category 1), for example, the background music of a bar is played, with the voice of the actor saying "Cheers!" and then the sound of the actor drinking in one gulp. In the case of the word "prank" (expression category 3), the voice actor approached

Fig. 2 Appearance of the recording session

the listener from behind and said "BOO! Did I scare you?" to surprise him/her. In the case of the word "maraud" (expression category 4), the sound effect of an approaching bike was played and the actor stated, "Argh! My bag was snatched!" In the case of the word "grotto" (expression category 6), the sound effect of dripping water in a cave was played along with a voice stating, "Wow! This is beautiful."

Prototype flashcard application

To apply this narration to English vocabulary learning, the proposed flashcard-type application was implemented using a tablet (Fig. 3). A flashcard is a card on which a word is written, depending on its purpose. There are many flashcard formats, such as illustrated and two-sided. To employ it for introducing English words, English and Japanese word pairs were displayed on the front side, while the narration was played.

The application was implemented with an iOS application (Flashcards Deluxe). The user can save words and narrations on his/her personal computer in advance and input correspondences between words and sounds using Excel. By synchronizing it with Dropbox, words and sounds are displayed on the tablet in a flashcard format. The user can thus learn the next word by swiping forward.

Experiment 1: Memorize under a controlled environment

To assess if the proposed method facilitates the memory retention of English words, we compared the number of words forgotten using the previous method (baseline) and that of the proposed method (proposed) (RQ1). The participants memorized a pair of words, that is, an English term and its Japanese translation by two methods. Both immediately and 1 week later, the user retrieved the memorized words in a memory test. To check the emotional state during memorization, we measured the SCR and compared it across the two methods (RQ2). The memorization method of English words may differ by each

Table 1 List of recorded narrations

Expression	Words (recorded length in seconds)
1. One person	Duress (7), peeve (13), barrow (9), feline (6), swig (7)
2. Two persons	Bawl (11), throes (7), bungle (10), schism (7)
3. Approach	Renege (8), pucker (7), aerate (9), cajole (8), navel (8), prank (3), suckle (7)
4. Touch	Maraud (8), pummel (6), rumple (7)
5. Circulate	Loony (14), hornet (7), heckle (8), fondle (9)
6. Together	Bemoan (11), honk (6), fondle (9), snoop (9), grotto (9), bovine (11), mirth (10), rapt (10)

Fig. 3 Appearance of the application

participant; hence, memorization time and the times a sound played are controlled in this experiment.

Participants

Since we are preparing to release this learning application to the public, we need to define the main target audience for the application. Since we plan to advertise through the voice actor's Twitter account, we decided to match the demographics of the experiment participants with the demographics of the voice actor's Twitter fans. Twitter users who tweeted the voice actor's name were approximately 60% male, and approximately 70% were in their 20s and 30s (using Mieruka analysis tool, Plus Alpha Consulting Co., Ltd.). Thus, we recruited participants who were men in their 20s and 30s.

We conducted experiments with 30 participants. They were hired and compensated through a temporary agency. The participants were all male native Japanese speakers ranging from 20 to 39 years old ($M = 27.6$, SD $= 6.1$). Fifteen of the participants had good English knowledge, with Test of English for International Communication (TOEIC) scores above 730 points (A or B level on the TOEIC proficiency scale). The remaining subjects were native Japanese speakers from a relatively lower socioeconomic bracket, who self-declared that they were not adept at memorizing English words. Two out of the 30 participants were excluded from the result as they reviewed the words after the test on the first day.

Vocabulary and materials

To know the tendency of forgetting rate of the word that is unknown to the participants, we selected ten nouns (loony, mirth, navel, prank, barrow, grotto, throes, bovine, hornet, honk) and ten verbs (bawl, maraud, pummel, aerate, rumple, heckle, cajole, swig, suckle, snoop), as shown in Table 1. These words were not included in the most frequently occurring 9000 words of the British National Corpus, with which the most Japanese may not learn this vocabulary level in his/her education; thus, participants may have been familiar. The average number of characters was 5.5 (four letters: three words; five letters: five words; six letters: twelve words), as experiments in a previous study which verify memory retention effect of learned words used ten target words of five or six letters. We used the Japanese translation of the words by the Weblio dictionary (Weblio 2005).

The average length of the narration of the proposed was 8 s (SD = 2 s). The sound of the baseline utilized the voice of an online dictionary (Goo(Goo dictionary 1999); Nippon Telegraph and Telephone resonant). The length of the voice recording was 0.8 s on average (SD = 0.2 s). In comparing forehand the quality of the sound for the four online dictionaries (Goo (Goo dictionary 1999), American Heritage (American Heritage Dictionary. Online dictionary 1969), Weblio (Weblio 2005), and Oxford Advanced Learner's (Oxford Learner's Dictionaries 1948)), Goo had the sound quality closest to the proposed. The sound quality of the proposed was based on a 44.1 kHz sampling rate, 24 bit depth, and two channels, whereas that of the existing method was based on a 44.1 kHz sampling rate, 32 bit depth, and two channels. Note that the experimenter adjusted the sound volume in advance using the tablet setting to mitigate the difference in the volume of both conditions.

Experimental design and procedure

The experimental plan involved within-participants. The same participants experienced both the baseline and proposed methods, and we compared the number of words forgotten among the methods. The experimental procedure is shown in Fig. 4. The participants first learned ten words under baseline and immediately performed a memorization test. The next ten words were learned using the proposed method, and the participants immediately took a memorization test as well. To eliminate the influence of the order effect, we also reversed the order of words to memorize under the baseline after the proposed method. To eliminate the influence of word memorability, words were assigned in random order. Additionally, we set a pre-test phase for understanding the prior knowledge of the participants and a practice phase for eliminating the influence of the practice effect.

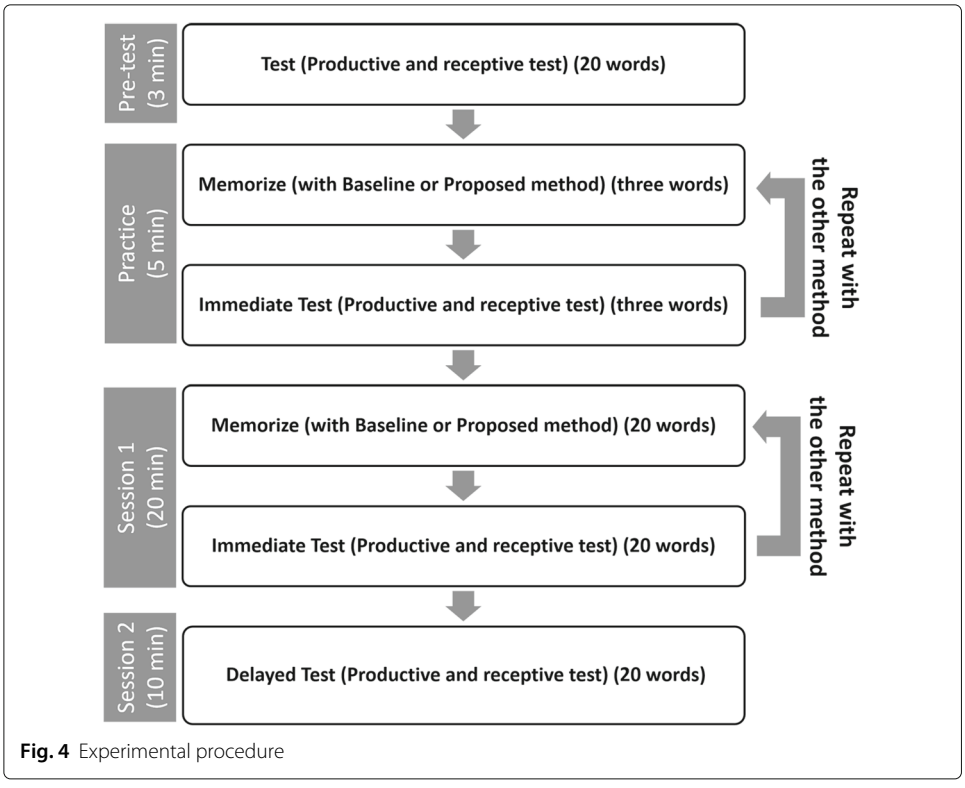

Fig. 4 Experimental procedure

The learning time per word was set to 30 s; hence, the total time of each memorization (previous and proposed) was 300 s (30 s $\times$10 words). The sound was played only one time, when the word was displayed on the tablet screen.

Pre-test

To understand prior knowledge of a word, a test was first conducted. It consisted of a productive and a receptive test. The productive test enabled the participants to enter the corresponding English term after the displayed Japanese term, while the receptive test was the reverse. However, in the productive test, to prevent the participants from answering with another semantically similar words, we showed several letters of the term and the correct answer for another letter as a hint. We tested the knowledge of all 20 words in random order.

Practice

To become accustomed with both methods, the participants practice them each with six words. The six words did not include words from the actual session. The first three words were learned by the baseline method; the next three words were learned by the proposed method. We requested the following of the participants: "Please actually remember English words with two kinds of voices: one is a normal voice, and the other is a voice actor's voice. The voice actor's narration is comprised of approximately 8 s of audio content for you to experience the meaning of the word. To control the approach for remembering the meaning, please do not use memorization techniques, such as the method of loci, equivoque, and other memory strategies. There is a learning time of 30 s for each word. Please remember it while repeating it."

Session 1

The participants first memorized ten English words by the baseline or proposed methods. Each participant was seated and engaged in memorization using a tablet (iPad Pro, Apple) and headphones (EAH-T-700, Technics) (Fig. 5 (left)). An English word and its Japanese

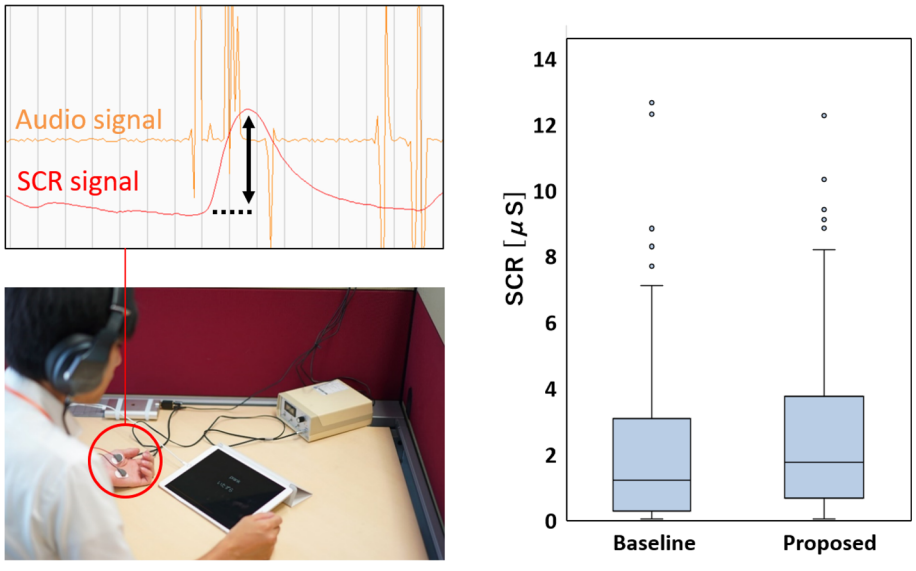

Fig. 5 Left: Electrodes attached to the participants' palms and sample SCR data. Right: SCR value of the baseline and proposed methods

translation were displayed for 30 s on the tablet screen, and the participants performed memorization while viewing the screen. Additionally, the sound was played in synchronization with the screen display. By swiping the touch screen, the next word appeared. After the ten-word memorization by the baseline method, a memory test was performed. The memory test consisted of a five-word test involving the translation of Japanese words into English and a five-word test of translating Japanese words into English. After that, using the proposed method, the user performed memorization, and then a memory test was conducted for the ten words by the same procedure. Considering the influence of the order effect, we also prepared a group in the reverse order for memorization by the baseline method after the proposed one. All participants used the same words. However, the word presentation order and sound quality (baseline or proposed) differed for each participant.

Session 2

One week after this experiment, only the memory test was conducted again. We subtracted the number of correct answers after one week from the number of correct answers immediately after the memorization of English words for the "forgetting rate." We then compared the number of forgotten items by the baseline and proposed methods. The procedure of the memory test was the same as in the previous stage.

The experiment was conducted over 2 days. The time required for the first day was 28 min and that required for the second day was 15 min. We did not explain to the participants that the same English word memory test would be carried out on the second day, we only explained that the same experiment would be carried out on the second day. This prevented the participants from perceiving the intention of the experiment and reviewing the memorized content between the experiment of the first day and the second day test.

Psychological response measuring and words scoring

During the experiment, electrodes were attached to the participants' palms to measure the SCR, which was calculated as the difference between the maximum and minimum values during the playing of a sound (Fig. 5 (left)). At the beginning of the experiment, the experimenter instructed the participants not to move their bodies while memorizing to ensure that no SCR artifact was included in the measurement data.

To score the responses, a sensitive scoring protocol was adopted (Barcroft 2002). In this method, if one character was correct, or if there were correct characters of 25% or more and less than 50% (the position may not be correct), it was set to 0.25 points. Similarly, if 25% or more and less than 50% of letters were correct, or if there were 50% or more and less than 75% of correct characters, it was set to 0.5 points. When 50% or more and less than 100% of characters were correct, or when an extra letter was added, it was set to 0.75 points.

Results and discussion

Figure 5 (right) shows the box plot of the SCR values for the baseline and proposed methods. The sample number from the baseline method was 280, with a total of ten measurements from 28 participants. On the other hand, the sample number of the proposed method was 279, where one sample data could not be measured due to hardware malfunction of the SCR measurement device. To validate data normality, we used the

Shapiro-Wilk normality test. As a result, the p value for both the baseline and proposed methods were $p < .05$, and the data were not normally distributed. To compare differences between the baseline and proposed methods, we used the Mann–Whitney test. As shown in Table 2, the mean rank SCR value of the proposed method (304.2) was significantly higher than the SCR value of the baseline method (255.9): $U = 32313.5$, $p < .05$, $r = 0.21$. This result suggests that the proposed method induced higher emotional arousal than the baseline method (RQ2).

Figure 6 (left) shows that under the baseline approach, the average number of correct answers decreased from 6.0 words to 2.4 words in 1 week, that is, the average forgetting rate was 3.6 words. On the other hand, in the proposed method, the average number of correct answers decreased from 5.5 words to 2.7 words in 1 week, that is, the average forgetting rate was 2.8 words.

To compare the forgetting rate between the baseline and proposed methods, their respective scores were subtracted from one another for each subject (Fig. 6 (right)). To validate data normality, we used the Shapiro-Wilk normality test, which indicated that data for both the baseline ($p = .80$) and the proposal ($p = .31$) were normally distributed, and we used a paired two-tailed t test for the statistical analyses. The participants' forgetting rate score for proposed method (M = 0.28, SD = 0.18) was significantly lower than that of the baseline method ($M = 0.35$, SD = 0.19): $t[27] = 2.36$, $p < .05$, $d = 0.41$. This suggested that the participants were able to memorize and retain more English language words with the proposed method than the baseline method (RQ2).

According to a feedback form completed after the delayed test, many positive comments were communicated by the participants: "The retention rate is remarkably different between the actor's voice and the conventional voice"; "I remember the story from the beginning to the end, and I can reproduce the story"; "The actor's voice is remarkably impressive compared to the conventional one, but it might be difficult to completely memorize English word pairs with one memorization"; "I would like to use it for learning English words if the application is released"; "I remembered vividly that the sound source moved from the left to the right of the headphones, and I also remember I felt a chilling sensation at this point."

Although our results suggested that the forgetting rate with the proposed method was lower than that with the conventional learning method, there remain some problems that must be addressed in future works. For example, the delayed test score only increased from about 20% for the baseline to about 30% for the proposed method. Most participants responded that the story was clearly remembered. On the other hand, there were some cases in which the Japanese recording was erroneous. For example, the word "loony" was recorded by the voice actor as a "crazy person" circling around the listener. Some subjects associated this meaning with "fear" based on the atmosphere in the recording. Additionally, the word "honk" contained a sound effect of a car approaching while a horn was sounded, and the voice of the actor cried "danger." One participant remembered the scream rather than the sound of the horn, and thus associated the meaning with "shout."

Table 2 The results of the Mann–Whitney test

Method	*N*	Mean rank	Mann–Whitney *U*	*p*
Baseline	280	255.9		
Proposed	279	304.2	32313.5	$p < .05$

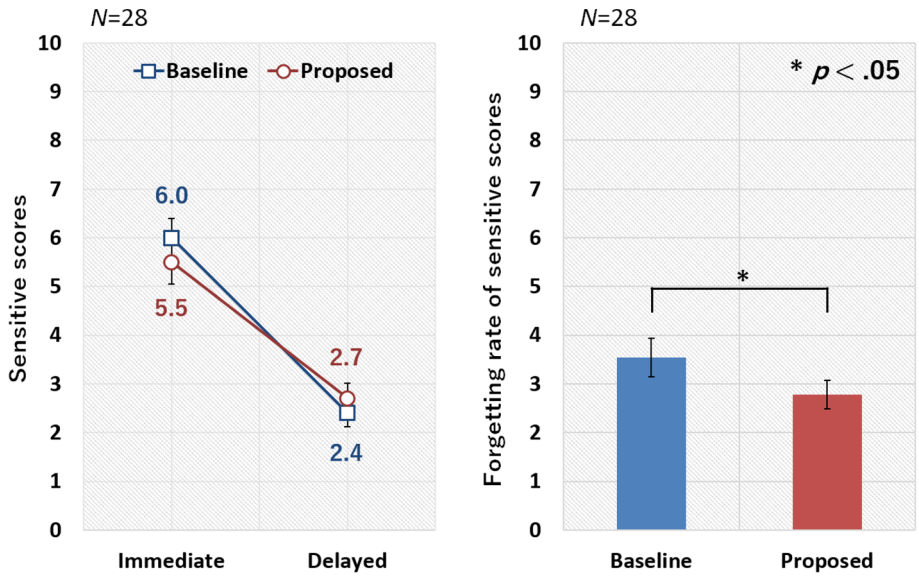

Fig. 6 Left: Average of the sensitivity scores in the immediate and delayed tests. Right: Forgetting rate of this score under the baseline and proposed methods. The error bar means standard error (SE)

It seems that a term that has multiple Japanese translations from a story is sometimes remembered in English as having a different meaning.

Additionally, although stories of English words and Japanese translations are easy to memorize, the connection with the spelling is often forgotten. Further, as mentioned by a participant, "Although the story and Japanese translation remain, the connection with English words often disappeared." "For the word 'cajole', the impression of the story was too strong, and the memory of the spelling disappeared." "I could remember the story, but I could not recall which word the story was associated with; if I heard it again I could learn it again and be able to make a firm connection."

Many learning experiences often use output and reflection to change the experience into knowledge (WEARABLE LANGUAGE TEACHER ELI 2017). However, in this experiment, the narration was listened to only one time with no opportunity to repeat it, making it somewhat different from the general process of experience learning. Therefore, it can be considered that a sufficient learning effect was not observed.

Experiment 2: Memorize under a freeform learning environment

To evaluate the memory retention effect of our method under freeform learning, the SCR measuring device was removed in this experiment. The learning time per word and total learning time were not controlled, and the participants could return to a previous word or skip a word by swiping on the tablet screen. Furthermore, the participants could play a sound any number of times. In this way, the participants memorized the word using the tablet application until they were satisfied (Table 3).

The learning process was also changed in this experiment. Specifically, the immediate review phase of converting experience into knowledge was added (Fig. 7). During the review phase, the participants responded to multiple-choice questions to memorize words.

Table 3 Points of modification from experiment 1

Points of modification	Experiment 1	Experiment 2
Immediate review	Without	With
Measuring device	Attached	Not attached
Learning time per word	30 s	Not controlled
Sound	One time	Not controlled
Can return or not	Cannot	Can return

Experimental design and procedure

We conducted the experiment with 30 participants. Two of 30 participants were excluded from the results because we could not obtain correct answers from them due to print errors on the test sheet. The remaining 28 participants were all male, native-Japanese speakers from 20 to 37 years old ($M = 25.7$, SD $= 4.7$) who were not included in Experiment 1. Similar to experiment 1, all participants were men in their 20s and 30s (see "Participants" subsection in the "Experiment 1: Memorize under a controlled environment" section). All participants were either university students or graduates. Their TOEIC scores were below 860 points. The words and sounds were the same as in experiment 1.

Similar to experiment 1, the experimental plan involved within-participants. The same participant experienced both the baseline and proposed methods, and we compared the number of words forgotten among the methods. The experimental procedure is shown in Fig. 8.

The participants first learned ten words by the baseline approach and then performed an immediate review for converting the experience into knowledge. After that, they performed an immediate retrieval test. The next ten words were learned using the proposed method. To eliminate the influence of the order effect, the word memorization order was reversed between the baseline and proposed methods. To eliminate the influence of word memorability, the words were assigned in random order. Additionally, we set a pre-test phase for understanding the prior knowledge of the participants and a practice phase for eliminating the influence of the practice effect.

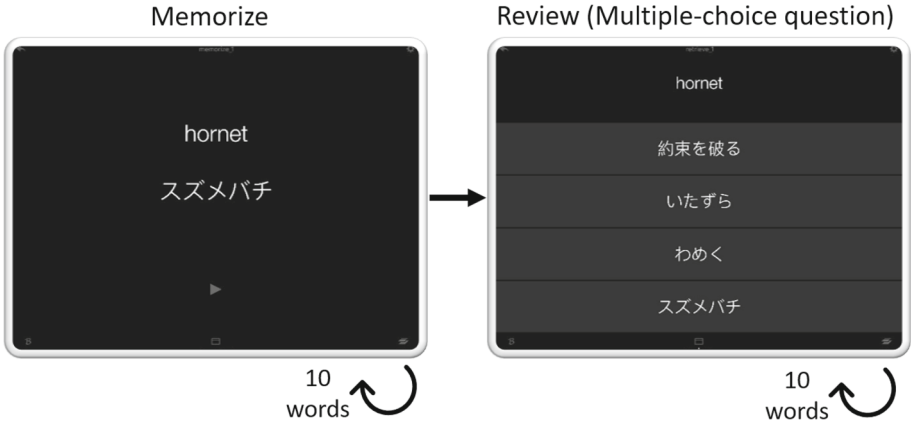

Fig. 7 Screen-shots from the memorize and review processes. By adding the review process, the participants may be able to convert their experience into knowledge

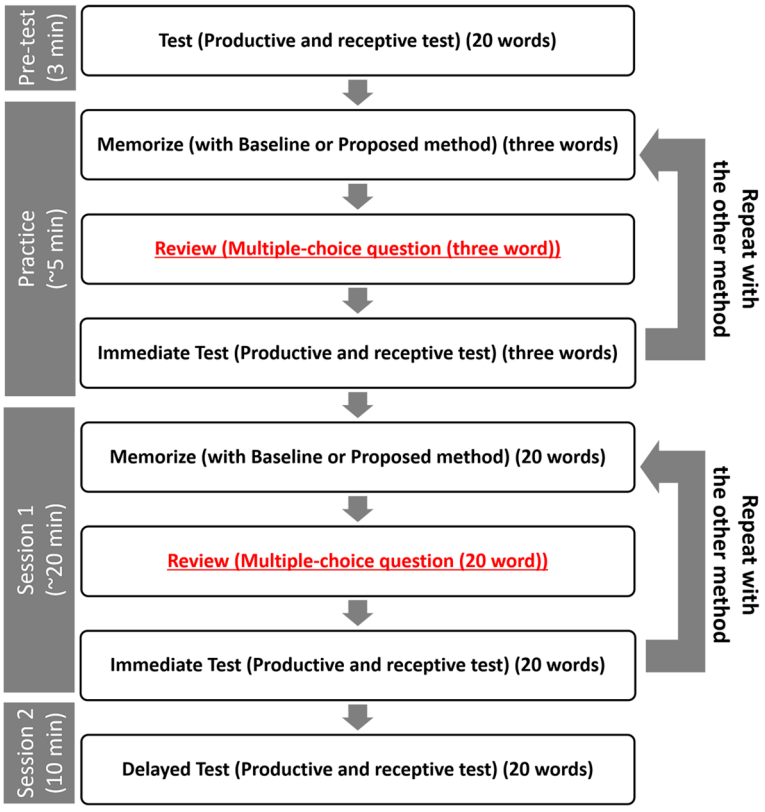

Fig. 8 Experimental procedure. The red underlined text is the additional process

Results and discussion

Figure 9 (left) shows that, under the baseline approach, the average number of correct answers decreased from 8.9 to 3.0 words in 1 week, that is, the average forgetting rate was 5.9 words. On the other hand, under the proposed method, the average number of correct answers decreased from 9.0 to 4.8 words in 1 week, that is, the average forgetting rate was 4.2 words. The correct answer rate after one week under the proposed approach increased to about 50% in this experiment compared to 30% in experiment 1. Under both methods, the scores of the immediate test were higher than those in experiment 1, which suggests that the participants could better memorize words in this experimental environment.

To compare the forgetting rate under the baseline and proposed methods, their respective scores were subtracted from one another for each subject (Fig. 9 (right)). To validate data normality, we used the Shapiro-Wilk normality test, which indicated that data for both the baseline ($p = .98$) and the proposal ($p = .99$) were normally distributed, and we used a paired two-tailed t test for the statistical analyses. The participants' forgetting rate score for proposed method ($M = 0.42$, SD $= 0.22$) was significantly lower than that of the baseline method ($M = 0.59$, SD $= 0.18$): $t[27] = 5.28$, $p < .05$, $d = 0.86$. This suggested that the participants were able to memorize and retain more English language words with the proposed method than the baseline method in the freeform learning environment.

We observed various participant behaviors during the experiment. Some participants memorized the words by continuously tapping the sound button while others by spelling the words with their finger on a table surface. Furthermore, after memorizing ten words,

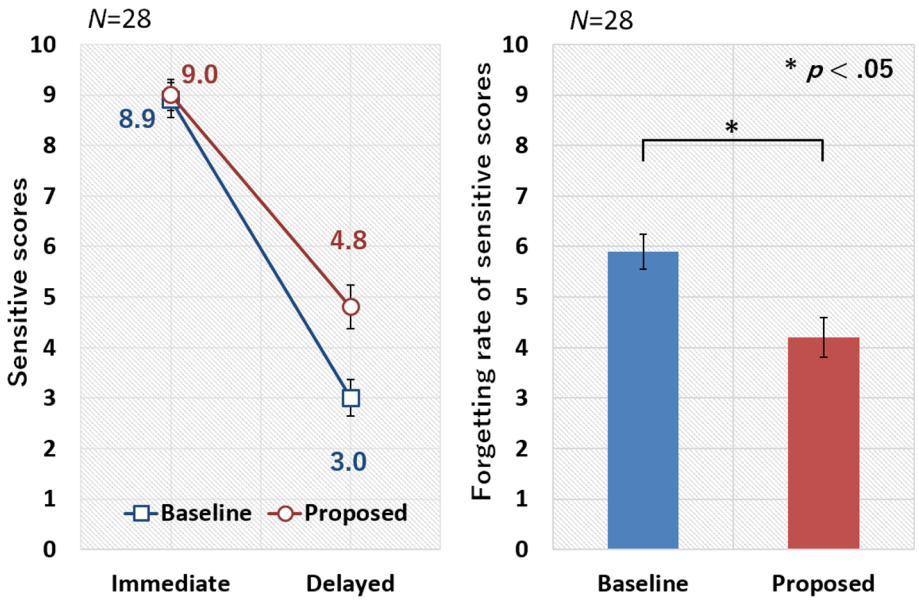

Fig. 9 Left: Average sensitive scores for the immediate and delayed tests. Right: Forgetting rate of this score under the baseline and proposed methods. The error bar means standard error (SE)

the participants usually checked their understanding by occluding the Japanese answer by hand. We believe that through this process, their experiences were converted to vocabulary knowledge. Consequently, the average review time for the ten English words was short: 63 s.

In the feedback form completed after the delayed test, many positive comments were communicated by the participants: "The voice with a story was easier for imagining the words than the normal one, thus the words in the actor's voice remained in memory"; "A stimulating story is easier to remember"; "In case of the actor's voice, I felt that the voice gives me a chance to remember even if I forget the word"; "Without a narration, I have to learn continuously. With a narration, it was fairly easy to remember, and learning efficiency was good."

Although the number of forgotten words decreased compared to experiment 1, some participants still forgot the spelling of the English words similar to experiment 1. The memory enhancement effect by emotional stimulation does not promote all the sensory information experienced, but may affect inhibition (arousal-biased competition theory) (Mather and Sutherland 2011). This suppression involves temporal (Strange et al. 2003) and spatial suppression (Kensinger et al. 2007), which means that the memory of the sensory information before and after the emotional stimulation and the surrounding information of the emotional stimulus are inhibited.

Additionally, attention is related to this inhibition and the spatiotemporal peripheral information is related to attention, which tends to not be remembered. Therefore, a scene where the emotion of the story is high expresses the meaning of the word, but it is still necessary to embed the information on spelling in that scene. Further, the attention may be targeting the auditory information, meaning the visual information presenting the spelling is suppressed. It is thus necessary to consider a visual presentation that calls attention to the spelling as well.

Table 4 shows the time length of the memorization phase, which was measured by an experimenter with timer application using a PC. The first and second lines are the times of experiment 1, and the third and fourth are the those of experiment 2. In experiment 2, the time length for proposed method ($M = 301$, SD $= 76$) was significantly longer than that of the baseline method ($M = 237$, SD $= 116$): $t[27] = 3.91$, $p < .05$, $d = 0.66$. In experiment 1, the time length of the memorization phase of all participants are controlled at 300 s, so statistically the result could not be calculated. However, the length of the baseline condition seems shorter than that of experiment 1. On the other hand, the length of the proposed condition is almost the same as that of experiment 1. The duration is almost the same, but the delayed score is improved; hence, by adding the immediate review and changing the learning environment, the proposed method could improve the delayed score.

Conclusion and future work

In this paper, we proposed a voice-enhanced emotional flashcard application for mobile phones through which a learner can perceive the meaning of English words. Emotional binaural voice narrations were used to enhance L2 vocabulary learning. Comparing the memory enhancement effects of the voice in the proposed method with a typical voice in the baseline approach, it was found that learning by the proposed voice makes it significantly easier to remember the English word and its translation. However, the spelling was still not memorized. We believe this relates to two aspects: content design theory and arousal-biased competition theory. In future works, we will employ audio content design theory and an attention induction method to reinforce memory retention. Additionally, we will expand the content of the proposed method. Furthermore, we will evaluate this method in an actual learning situation.

In addition, we conducted two experiments with only men in their 20s and 30s as they were identified as the main target user group for our application. However, expanding the participant demographics and generalizing the results to women is also important and should be addressed in future works. The study of IADS (Bradley and Lang 2007) suggested that emotional reactions induced by a sound differ between males and females. Therefore, our results reported here may not be generalizable to women. However, when women experienced this application at a closed event, none of them had a negative response to the voices or applications. Moreover, similar to the experiment participants, they showed emotional reactions of arousal. Therefore, we are planning a controlled experiment with only female participants as a future work.

Finally, this paper suggests that emotive story-based binaural narration promotes the memory retention of English words. However, it is not clear which element of this narration contributed to the results. For example, the two experiments conducted in this study

Table 4 Time length in seconds of the memorization phase in experiments 1 and 2

Experiment	Condition	Mean time	SD
Experiment 1	Baseline	300	–
	Proposed	300	–
Experiment 2	Baseline	237	116
	Proposed	301	76

The time length of experiment 1 is controlled at 300 s

include two independent variables, namely an emotive narration versus a non-emotive narration and a story-based narration versus a non-story-based narration. As such, further investigation to better understand how individual factors affect memory retention is needed.

Abbreviations
Not applicable.

Acknowledgments
We would like to thank M. Kimura and T. Wakamiya for their suggestions with respect to language education. We also extend our thanks to S. Oguni and H. Takumi for their cooperation during the brainstorming and recording process. Finally, we are grateful to A. Hautasaari for carefully proofreading the manuscript. This research was (partially) supported by JST PRESTO (Grant No. JPMJPR1658).

Authors' contributions
SF contributed to all aspect of this manuscript: conception and design of the study, analysis and interpretation of data, collection and assembly of data, drafting of the article, critical revision of the article for important intellectual content, and final approval of the article. The author read and approved the final manuscript.

Authors' information
I was born in 1986 and received a Ph.D. in engineering from the University of Electro-Communications in 2013. I was a visiting student of Camera Culture Group at MIT Media Lab supported by "Japan Society for Promotion of Science (JSPS) Research Fellowships for Research Abroad," and a project researcher of Graduate School of Information Science and Technology at the University of Tokyo. I am currently an Assistant Professor of Graduate School of Information Science and Technology at the University of Tokyo and a researcher at Japan Science and Technology Agency (JST) PRESTO. My research interests are IA (Intelligence amplification), virtual reality, entertainment computing, and human emotions.

Funding
This research was (partially) supported by JST PRESTO (Grant No. JPMJPR1658).

Availability of data and materials
Not applicable.

Competing interests
Not applicable.

Received: 18 April 2019 Accepted: 23 September 2019
Published online: 08 November 2019

References
Nation, I.S.P. (2006). How large a vocabulary is needed for reading and listening? *500*, 59–82. https://doi.org/10.3138/cmlr.63.1.59.
van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: the same or different from reading comprehension? *Applied Linguistics*, *34*(4), 457–479. https://doi.org/10.1093/applin/ams074.
Hwang, G.-J., & Fu, Q.-K. (2019). Trends in the research design and application of mobile language learning: a review of 2007–2016 publications in selected SSCI journals. *Interactive Learning Environments*, *27*(4), 567–581. https://doi.org/10.1080/10494820.2018.1486861.
Hwang, G.-J., & Wu, P.-H. (2014). Applications, impacts and trends of mobile technology-enhanced learning: a review of 2008-2012 publications in selected SSCI journals. *International Journal of Mobile Learning and Organisation*, *8*(2), 83–95. https://doi.org/10.1504/IJMLO.2014.062346. PMID: 62346. https://www.inderscienceonline.com/doi/abs/10.1504/IJMLO.2014.062346.
Weblio (2005). *Online dictionary*: Weblio, Inc. http://www.weblio.jp/.
Goo dictionary (1999). Online dictionary. NTT Resonant Incorporated. https://dictionary.goo.ne.jp/.
Jayme Adelson-Goldstein, N.S. (2015). *Oxford Picture Dictionary Monolingual (American English) Dictionary for Teenage and Adult Students (Oxford Picture Dictionary Second Edition)*. Oxford: Oxford University Press.
American Heritage Dictionary. Online dictionary (1969). Houghton Mifflin. https://ahdictionary.com/.
mikan (2014). *English learning application*: mikan Co., Ltd. http://mikan.link/.
Smart Language Apps Limited (2015). *Learn English (US) Flashcards, English learning application*: Smart Language Apps Limited. https://itunes.apple.com/us/app/learn-english-usflashcards/id970002864?mt=8.
Wright, S., Fugett, A., Caputa, F. (2013). Using e-readers and internet resources to support comprehension. *Educational Technology & Society*, *16*(1), 367–379.
Moetan, DS (2008). *FACTORY Co.*: IDEA, Ltd. http://www.ink-chan.com/others.html.
Hiroshi, O. (2014). *Moesta Moerutodaieigojyuku*: NOISE FACTORY co.,ltd. https://web.archive.org/web/20080730151146/http://moe-sta.jp/.
Ogura H. (2014). *Maruoboe Eitango 2600*: KADOKAWA CORPORATION.
Hung, H.-T., Yang, J.C., Hwang, G.-J., Chu, H.-C., Wang, C.-C. (2018). A scoping review of research on digital game-based language learning. *Computers and Education*, *126*, 89–104. https://doi.org/10.1016/j.compedu.2018.07.001.
Kensinger, E.A., & Corkin, S. (2003). Memory enhancement for emotional words: are emotional words more vividly remembered than neutral words? *Memory & Cognition*, *31*(8), 1169–1180. https://doi.org/10.3758/BF03195800.

Kleinsmith, L.J., & Kaplan, S. (1964). Interaction of arousal and recall interval in nonsense syllable paired-associate learning. *Journal of Experimental Psychology*, *67*(2), 124–126. https://doi.org/10.1037/h0045203.
McGaugh, J.L. (2003). *Memory and emotion: the making of lasting memories*. New York City: Columbia University Press.
Phelps, E.A., LaBar, K.S., Spencer, D.D. (1997). Memory for emotional words following unilateral temporal lobectomy. *Brain and Cognition*, *35*(1), 85–109. https://doi.org/10.1006/brcg.1997.0929.
Cull, W.L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology*, *14*(3), 215–235.
Russell, J.A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178.
Hamann, S.B., Ely, T.D., Grafton, S.T., Kilts, C.D. (1999). Amygdala activity related to enhanced memory for pleasant and aversive stimuli. *Nature Neuroscience*, *2*(3), 289–293. https://doi.org/10.1038/6404.
LaBar, K.S., & Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience*, *7*(1), 54–64. https://doi.org/10.1038/nrn1825.
Coles, K., & Tomporowski, P.D. (2008). Effects of acute exercise on executive processing, short-term and long-term memory. *Journal of Sports Sciences*, *26*(3), 333–344. https://doi.org/10.1080/02640410701591417. PMID: 18074301. https://doi.org/10.1080/02640410701591417.
Bradley, M.M., & Lang, P.J. (2007). The International Affective Digitized Sounds (; IADS-2): affective ratings of sounds and instruction manual. Tech. Rep. B-3.
Lang, M.M.B., Peter, J., Cuthbert, B.N. (2008). International Affective Picture System (IAPS): affective ratings of pictures and instruction manual. Technical Report A-8. University of Florida.
3D sound attraction of Joypolis (2016). CA Sega Joypolis Ltd. Retrievedfromhttp://tokyo-joypolis.com/language/english/attraction/3rd/hozuki.html.
Barratt, E.L., & Davis, N.J. (2015). Autonomous sensory meridian response (asmr): a flow-like mental state. *PeerJ*, *3*, 851.
Suzuki, Y. (2000). *DUO 3.0*. Tokyo: ICP Inc.
Cavus, N., & Ibrahim, D. (2009). m-Learning: an experiment in using SMS to support learning new English language words. *British Journal of Educational Technology*, *40*(1), 78–91. https://doi.org/10.1111/j.1467-8535.2007.00801.x.
Gassler, G., Hug, T., Glahn, C. (2004). Integrated Micro Learning - an outline of the basic method and first results. *Interactive Computer Aided Learning*, 1–7.
Nakata, T. (2015). Effects of expanding and equal spacing on second language vocabulary learning. *Studies in Second Language Acquisition*, *37*(04), 677–711. https://doi.org/10.1017/S0272263114000825.
Luis, A., & von Severin, H. (2011). Duolingo. Retrievedfromhttps://www.duolingo.com/.
Nessel, D.D., & Dixon, C.N. (2008). *Using the language experience approach with English language learners : strategies for engaging students and developing literacy*, (p. 171). Thousand Oaks: Corwin Press.
WEARABLE LANGUAGE TEACHER ELI (2017). monom, 1-10 Group, and HAKUHODO PRODUCT'S. Retrievedfromhttp://eli-talk.com/en/.
Al-Mekhlafi, K., Hu, X., Zheng, Z. (2009). An approach to context-aware mobile Chinese language learning for foreign students, In *2009 Eighth International Conference on Mobile Business*. https://doi.org/10.1109/ICMB.2009.65 (pp. 340–346).
Dearman, D., & Truong, K. (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*. https://doi.org/10.1145/2207676.2208598 (pp. 1391–1400).
Hsieh, H., Chen, C., Hong, C. (2007). Context-aware ubiquitous English learning in a campus environment, In *Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007)*. https://doi.org/10.1109/ICALT.2007.106 (pp. 351–353).
Ogata, H., & Yano, Y. (2004). Context-aware support for computer-supported ubiquitous learning, In *The 2nd IEEE International Workshop on Wireless and Mobile Technologies in Education, 2004. Proceedings*. https://doi.org/10.1109/WMTE.2004.1281330 (pp. 27–34).
Zhu, Y., Wang, Y., Yu, C., Shi, S., Zhang, Y., He, S., Zhao, P., Ma, X., Shi, Y. (2017). ViVo: Video-Augmented Dictionary for Vocabulary Learning, In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*. https://doi.org/10.1145/3025453.3025779. http://dl.acm.org/citation.cfm?doid=3025453.3025779 (pp. 5568–5579). New York: ACM Press.
Loftus, E.F. (1996). *Eyewitness Testimony*. Oxford: Oxford University Press.
Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied linguistics*, *28*(1), 46–65.
Vetrugno, R., Liguori, R., Cortelli, P., Montagna, P. (2003). Sympathetic skin response. *Clinical Autonomic Research*, *13*(4), 256–270. https://doi.org/10.1007/s10286-003-0107-5.
Oxford Learner's Dictionaries (1948). *Online dictionary*: Oxford University Press. Retrievedfromhttp://www.oxfordlearnersdictionaries.com/.
Barcroft, J. (2002). Semantic and structural elaboration in L2 lexical acquisition. *Language Learning*, *52*(2), 323–363. https://doi.org/10.1111/0023-8333.00186.
Mather, M., & Sutherland, M.R. (2011). Arousal-Biased Competition in Perception and Memory. *Perspectives on psychological science : a journal of the Association for Psychological Science*, *6*(2), 114–33. https://doi.org/10.1177/1745691611400234.
Strange, B.A., Hurlemann, R., Dolan, R.J. (2003). An emotion-induced retrograde amnesia in humans is amygdala- and beta-adrenergic-dependent,. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(23), 13626–31. https://doi.org/10.1073/pnas.1635116100.
Kensinger, E.A., Garoff-Eaton, R.J., Schacter, D.L. (2007). Effects of emotion on memory specificity: memory trade-offs elicited by negative visually arousing stimuli. *Journal of Memory and Language*, *56*(4), 575–591. https://doi.org/10.1016/j.jml.2006.05.004.

Publisher's Note