Iyer, S. et al. (Eds.) (2022). Proceedings of the 30th International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Effects of Binaural Audio on English Vocabulary Learning

Kosuke SHIMIZU[a*], Shogo FUKUSHIMA[b] & Takeshi NAEMURA[c]
[a]*Yamagata Prefectural Touohgakkan Senior High School, Japan*
[b]*Kyushu University, Japan*
[c]*The University of Tokyo, Japan*
*kosuke05816@icloud.com

Abstract: The English Picture Dictionary is a dictionary of stock pictures that represent meanings of English words. Following this, our aim was to create an audio repository of stock sounds that similarly represent meanings of English words. To effectively express the meaning of various English words through sound, we used binaural recording technology. For example, when recording the word "fetch," it was represented as a sound movement by moving a dummy microphone head from side to side, while pronouncing the word. This paper reports the results of a comparison between the effects of binaural and monaural audio on the learning of English vocabulary.

Keywords: Binaural recording, computer-assisted language learning, vocabulary learning

1. Introduction

Learning vocabulary is essential for acquiring a second language, such as English. According to the bilingual dual-coding theory (Allan & Desrochers, 1980), vocabulary is retained by forming an association between verbal and non-verbal information. For this reason, the Picture Dictionary (Adelson-Goldstein, 2015) and movies (Zhu et al., 2017) are commonly used in vocabulary learning. The Picture Dictionary is a stock picture dictionary that represents meanings of English words. By following similar lines, our aim was to create an audio version of stock sounds that represent meanings of English words.

With the increasing digitization of learning material, learning formats that use rich audio-visual multimedia content are becoming more common, and are consequently, replacing paper-based material. In second language learning, learning methods that employ smartphones (Dearman et al., 2012) and virtual spaces (Vazquez et al., 2018) are often proposed. EmoTan (Fukushima, 2019) is a learning system that specifically utilizes auditory technology; it employs emotional narration and stories in the learning of English words and utilizes binaural recording technology to create sounds through emotional expressions. Fukushima (2019) demonstrated how the use of EmoTan enhances retention of English words, but the experiment involved two variables that change simultaneously, that is, the presence or absence of an emotional story and binaural recording technology. It did not examine the effect of binaural recording technology alone.

Therefore, this paper focuses on the presence or absence of binaural recording and examines its effect on the learning of English vocabulary. For this, we first categorized English words according to whether they could be represented by binaural recording technology or not. Next, 40 words were selected from the dataset, and their audio content was created in a binaural recording studio by a professional voice actress. Finally, a psychological experiment was conducted on these 40 words, to assess the effect of binaural recording on the retention of English words.

2. Word Selection and Binaural Audio Recording

First, it was necessary to select words whose meanings could be expressed through binaural recording technology. For this purpose, we used a cloud worker service (Amazon Mechanical Turk, Amazon.com Inc.) to classify the words. A total of 12,000 words were classified from the SVL word list (published by ALC PRESS INC), which were considered useful for Japanese learners. For each word, seven workers

assessed the suitability of the method. We also used the Common European Framework of Reference for Languages (CEFR), which provides a common basis for the explicit description of objectives, content, and methods for second/foreign language education, to check the vocabulary level of the classified English words. The following question items were designed for workers to judge the words accurately. "Can the meaning of the above word be conveyed using voice, sound effects, sounds made by objects/things, or a combination of these, without the word being spoken out loud? (Our goal is to realize stock sounds/narrations in a binaural recording studio, using the expressive voice of professional voice actors, sound effects, BGMs, and various objects/things)." In addition to this, we confirmed whether the words' prefixes and suffixes encouraged or discouraged the discerning of their meaning or memorization, and if they were at CEFR level B2 or higher, considering that our subjects were high school students.

Of the classified words, 40 English words were selected, for which more than half of the workers had answered "Yes," and binaural audio was recorded for them. Images from the recording are on the left of Figure 1. The motion information in the words was reproduced around a microphone with a dummy head. Some of the words and expressions are shown on the right of Figure 1.

Words	Audio expression
whirl	Circulate: pronounce while circling around the dummy head microphone.
unleash	From bottom to top: pronounce while moving from the bottom of the dummy head microphone to the top.
imminent	Approach: pronounce while approaching the dummy head microphone.
fetch	Left to Right: pronounce while moving to the left or right of the dummy head.
prairie	Facing the wall: pronounce while facing the wall.
aerate	Whispering: pronounce as if whispering near the dummy head microphone

Figure 1 Left: Recording with a dummy head, Right: Examples of words and their expressions

3. Experiment: Effect of Binaural Audio on Retention of English Words

To assess the effect of the presence or absence of binaural audio on retention of English words, we compared the number of words that could be memorized with monaural sound (baseline) vs. binaural sound (proposed). The participants memorized 20 English words for each audio, and attempted a test immediately after memorization, followed by another test a week later. In this manner, we could verify the effects with regard to two aspects: short-term and long-term memory retention. The experiment's detailed procedure and appearance are shown in Figure 2.

The experiment had a within-participants design, with all the participants participating in both conditions (baseline and proposed). There were a total of 20 participants. Of the 20 participants, 15 were male and 5 were female, of an average age of 16.3 years. We confirmed whether all of them had a CEFR score of A2 or lower.

The 40 English words selected were divided across two sets (Set1: blast, dizzy, ascend, flock, bawl, eclipse, crave, dismay, bemoan, mirth, whirl, loony, maraud, bovine, peeve, barrow, suckle, yarn, blast, and prank, Set2: shiver, basin, sniff, buzz, pummel, gust, cajole, sneeze, swig, grotto, implore, honk, snoop, navel, aerate, tumble, rumple, schism, tease, and hornet). We took care to ensure that the word sets did not affect the experiment's results. Specifically, we assigned the participants equally to all combinations of word sets (Set 1 and Set2) and voice conditions (baseline and proposed).

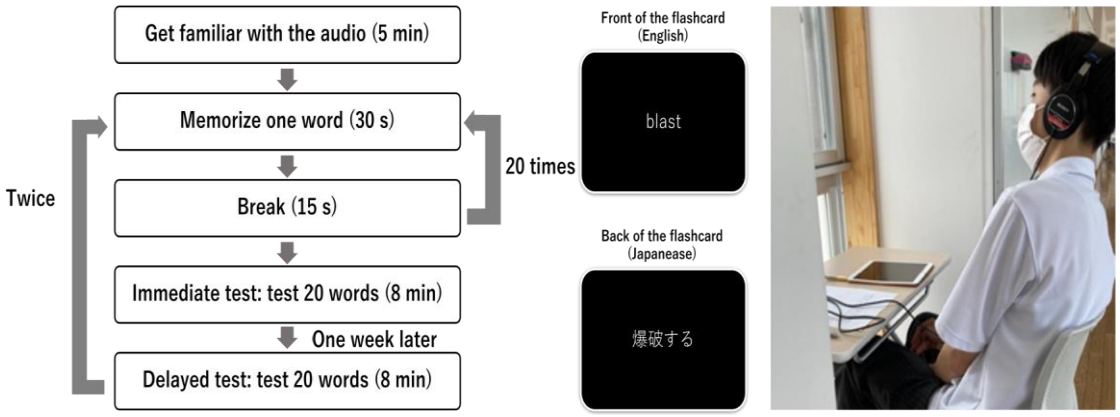

Figure 2 Left: Procedure of the experiment, Right: Image of a participant memorizing

4. Test scoring and the result

The vocabulary test included multiple-choice responses (left of Figure 3). The 20 English words learned were printed on a sheet, on which the Japanese translations were printed at the bottom. The order of the words was shuffled, and participants had to select the correct combination. The percentage of correct responses was calculated based on the number of words answered correctly.

On the right of Figure 3 are the mean scores and standard errors for the monoaural and binaural conditions. We conducted a two-factor ("Monoaural/Binaural" and "Immediately after/One week after") repeated measures ANOVA on the conditions, and the result showed a significant difference in the audio factor ($F(1, 19) = 8.40$, $p<.01$) and test factor ($F(1, 19) = 102.82$, $p<.01$). However, there were no significant differences in the interaction effects ($F(1, 19) = 3.14$, $p<.10$). The results suggest that the score for binaural is superior to that for monoaural, and the score for the test conducted immediately after is superior to that conducted one week after.

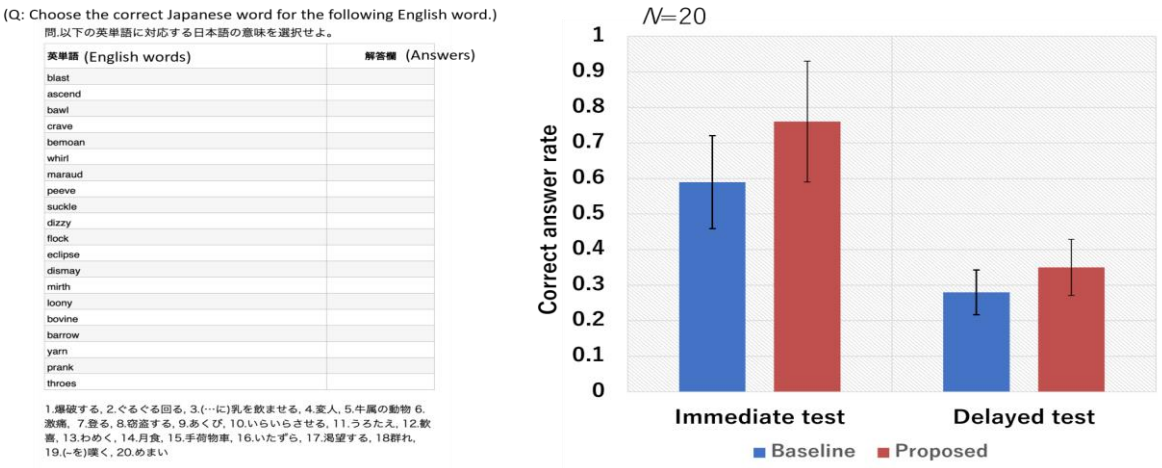

Figure 3 Left: Example of test screen; Right: Mean and standard error of test results

5. Conclusion and Future Work

This paper evaluated the effect of the presence or absence of binaural audio on the learning of English vocabulary. In the future, we shall evaluate the effect of learning over a longer period, as well as develop learning software.

Acknowledgements

This work is supported by UTokyoGSC and JST-Mirai Program Grant Number JPMJMI21D3, Japan.

References

Adelson-Goldstein, J., & Shapiro, N. (2015). *Oxford Picture Dictionary Monolingual (American English) Dictionary for Teenage and Adult Students (Second Edition)*. Oxford University Press.

Allan, P., & Desrochers, A. (1980). A dual-coding approach to bilingual memory. *Canadian Journal of Psychology/Revue canadienne de psychologie. 34* (4). 388.

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 (CHI '12*) (pp.1391–1400).

Fukushima, S. (2019). EmoTan: enhanced flashcards for second language vocabulary learning with emotional binaural narration. *Research and Practice in Technology Enhanced Learning*. *14* (1).

Vázquez, C., Xia, L., Aikawa, T., & Maes, P. (2018). Words in Motion: Kinesthetic Language Learning in Virtual Reality. In *2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)* (pp. 272–276).

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* (pp. 5568–5579).