Proceedings of the 3rd International Conference on Computing Innovation and Applied Physics
DOI: 10.54254/2753-8818/31/20241089

Design and implementation of English vocabulary orthographical revision application based on Matlab

Qi Cao

College of Automotive Engineering, Jilin University, Changchun, China

caoqi1520@mails.jlu.edu.cn

Abstract. Apropos of English learning, vocabulary learning is one of the most essential parts. Thanks to modern technology, recent years has witnessed scads of vocabulary learning apps springing in the app market which include greatly comprehensive functions. However, among those popular apps, some universal handicaps still exist in the design of the review section, for example, the absence of the freedom for the users to classify the cognitive or priority grade of vocabulary and the low effect of some reviewing methods that could not satisfy the need of the language learners to fully master the orthography of words. This paper aims at contriving an app serving as a supplementary part of English vocabulary learning in order to remedy those inadequacies of mainstream applications or other self-learning methods. In this paper, literature study, graphology, and mathematical modelling are involved. Based on Matlab GUI/App Designer, this paper presents a new method to review English words in which users could customize their lexicon, as well as define the priority of each word and practice spelling of vocabulary on an hourly-based system instead of a daily-based one that most of the existing online platforms adopted. Through creating a discrete probability density function for each word, this app could forward a random word in the glossary for users to practice orthography, obeying the probability density distribution that has a mathematical relationship with familiarity and priority of the words as well as the time factor (Ebbinghaus forgetting curve). Thus, users could thoroughly grasp the orthography of target words.

Keywords: Vocabulary learning, forgetting curve, mathematic model

1. Introduction

Since English is the most commonly used language in the world, English learning has always been a hot pedagogical issue all over the world and throughout history. Vocabulary learning is undoubtedly the pivot in systematic English study. Wilkins, a British scholar, has expressed his views on the importance of vocabulary in his book *Linguistics in Language Teaching*, claiming that ‘Without grammar very little can be conveyed; Without vocabulary nothing can be conveyed’ [1].

Nowadays, more and more learners begin to seek education through online platform depending on mobile devices. According to a survey targeting the generation after 00s in China, 65% of youngsters born after 00s would choose to use their cell phones as a learning tool. (Report on Internet learning behaviour of generation after 00s in China, 2016) [2]. Today, a rapid increment in both quality and quantity of theories and practices in the scientific domain of education and cognitive psychology, accompanied by the skyrocketing level of advanced technology, gives rise to a number of apps and online platforms featuring on English vocabulary learning. For instance, in Chinese app market, there are scads of vocabulary learning apps like BaiCiZhan (featuring on association-memorizing method), BuBei (featuring on establishing the connection between new words and famous English movie series) and Shanbei (featuring on powerful lexicon managing system). And an increasing number of English learners begin to invest time and money on those mobile applications. Hitherto, the downloads of BaiCiZhan have achieved 231 million in App Store China. However, there are still some common defects in the reviewing section among all vocabulary learning apps, for example, a lack of freedom in decide priority and importance of each word by users themselves and the absence of a more scientific mathematic model to review the learned words as well as the verbosity of some low efficiency methods of revision. And also, fragmented learning is becoming an inevitable tendency. In this paper, fragmented learning mainly refers to fragmented learning time. How the apps should adjust the learning content to suit the fragmented learning time is becoming an imperative matter. Based on the background hereinbefore, this paper intends for an app implementing a new method of reviewing the learned words in the means of Matlab GUI/App Designer, involving a self-adaptive mathematic model concerning the probability density of words which takes into consideration the effect of the forgetting curve, familiarity and priority of words.

This paper depends mainly on two theoretical model: Ebbinghaus forgetting curve model and L2 vocabulary acquisition model. Ebbinghaus forgetting curve serves as an irreplaceable role in the practices of English vocabulary learning. Learning English vocabulary with the application of forgetting curve enable learners' retention of fresh vocabulary to transform from a short-term one to a long-term one by increasing the intensity and frequency, solving the problem of ‘easy to forget’ [3]. L2 vocabulary acquisition model includes definitions and dimensions of words, mental representation styles of L2 vocabulary, L2 vocabulary knowledge framework, and the L2 Vocabulary Acquisition Process [4].

The paper is aiming at improving the efficiency of reviewing English vocabulary and ameliorating some amiss and superfluous reviewing methods through the establishment of a more scientific mathematic model and provide with the possibility of a new thread to conduct vocabulary learning.

2. Methodology

2.1. Adoption of orthographical exercises

Great advances have been achieved in mobile smart devices, which facilitating the advent of vocabulary learning apps. A few apps have integrated associative memory method into vocabulary learning. For instance, BaiCiZhan takes advantage of numerous pictures related to the English words to help memorize the Chinese meaning of the corresponding English word. However, Dunlosky claimed that the effects of using mental imagery to learn from text may be rather limited and not robust [5]. Moreover, most of the applications in current market implement rereading as the main revision methodology, while there are abundant researches indicating testing and practice are more effective than rereading in the formation of the long-term memory. Taking a memory test not only assesses what one knows, but also enhances later retention, a phenomenon known as the testing effect [6]. In his research, they compared the data of learning results from two groups of students (SSSS, STTT) where the students were required to learn from a certain article. Students of STTT group were asked to write down anything as much as they could remember after an only 5-min study, while students from SSSS group barely studied the article only through rereading, instead of doing practice. It turned out that the forgetting velocity of STTT group was largely less than that of SSSS group. Another research conducted by Rowland (2014) suggest that the tougher is the retrieving process, the firmer is the long-term memory formed [7]. Thus, the orthographical exercise, the best way to memorize the spelling of words with the lowest retrieval strength, is considered as one of the most effective ways in reviewing learned vocabulary.

And another requirement for the review system should be the assurance of great freedom to select the time for revision. There are two reasons standing for the preference of a system depend on discontinuous time and fragmented study in lieu of a time-continuous and time-consuming one. The prime reason is that more and more learners have difficulties sparing a long consecutive time learning vocabulary due

DOI: 10.54254/2753-8818/31/20241089

to the heavy load of study or work. Secondly, it is shown spacing or distributing study engages the very processes that support learning, comprehension, and remembering [8]. That means separating the time for learning into short sections will be helpful in the formation of long-term memory. It can be seen that the design of the app should feature on fragmented and discrete learning time, which should also ensure that the program should forward the optimum learning content at any time.

2.2. Establishment of mathematic model

2.2.1. Model of forgetting curve and refinement. Ebbinghaus forgetting curve delineates the natural law of memory. It discloses the mathematical relationship between memory loss and time (see Figure 1).

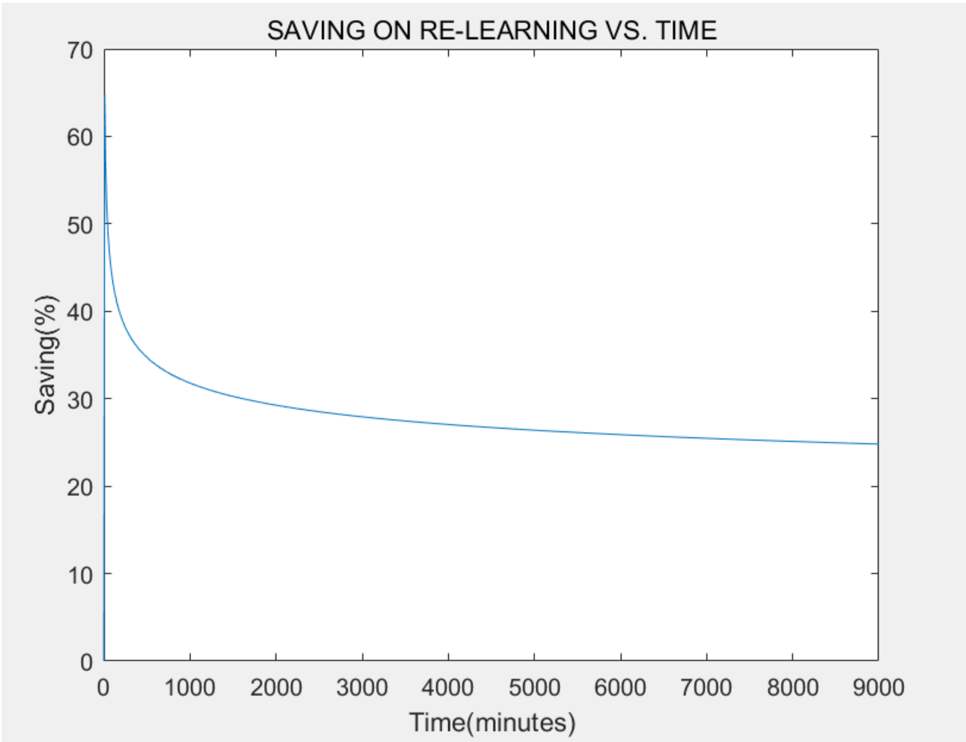

Figure 1. Forgetting curve adapted from Hermann Ebbinghaus (1885) [9]

In this model, the volume of memory is quantified, concatenating the domain of cognition and mathematics. The curve was incipiently fitted with Logarithmic Function [9]. However, the original model has several critical flaws and limitation:

1) The incipient experiment carried out by Ebbinghaus was only done under the situation where the testees tried to remember random permutations of certain characters in lieu of words with specific meanings.

2) The fitting method of the forgetting curve is timeworn and may not accurately describe the relationship between time and retention.

3) The curve only represents the relationship when memorizing for one time, not repetitive times.

In consideration of those factors, this paper adopts literature research to modify the memory curve which is applied in the mathematical model. In a research, Cai (2016) has proposed a self-adaptive model to make modification on the application of the forgetting curve [10]. Based on the research by Chris and Robert [11], where the conclusion is drawn that a power function could fit the forgetting curve more appropriately, Cai has introduced a methodology in which the parameter of the function could be adjusted automatically through feedback from respondents. The fitting power-law function proposed by Chris and Robert is as follow:

$$m = M \times \Delta t^{-\beta} \tag{1}$$

where m is retentive memory, dimensionless; M is Memory coefficient constant, dimensionless; Δt is time interval (unit: minute); β is Memory attenuation coefficient. In that self-adaptive model, β is considered as the controlled variable. They separate the feedback into 3 categories: unknown, ambiguous and conversant, with the corresponding value of β: 0.4307, 0.2038 and 0.1056.

2.2.2. Simplification of memory model. To simplify the model, M could be treated as 1 regardless of complexity of a certain word. The maximum of the membership function of memory is 1, meaning the retentive memory is 100%, which could be also interpreted as there's 100% probability recall the orthography of the word. In other words, a complete retention of certain words, while the minimum of the function is 0, meaning an absolute oblivion. Thus, the function of the relationship between memory and time could be figured out:

$$M(i) = [60 * \Delta t(i)]^{-\beta} \tag{2}$$

where M(i) is membership degree of retentive memory of word(i) in a certain time interval (unit of time: hour).

2.2.3. Establishment of discrete probability density function of occurrence for vocabulary. Two variables are defined to describe the familiarity and priority of each word: FV(i) and IMPT(i) (both dimensionless). It is assumed that original FV(i) is 100 and each time the user gives feedback FV(i) is changed according to whether the answer given by the respondent is correct or not. IMPT(i) is manually set by the respondent to decide the priority of each word (Scope: from 1 to 3, integer) and the higher is IMPT(i), the higher is the probability of occurrence of the corresponding word. These two properties describe the subjective factor of probability of occurrence, so that it is defined:

$$SF(i) = FV(i) * IMPT(i) \tag{3}$$

where SF(i) and M(i) respectively represents the two sections consisting the decider in the probability density function: subjective factor and objective factor. The discrete probability density function could be established:

$$P(i_0) = \frac{SF(i_0)^{1-M(i_0)}}{\sum_{i=1}^{i_{max}} (SF(i)^{1-M(i)})} \tag{4}$$

Where $P(i_0)$ is the probability of occurrence of the word NO. i_0.

2.2.4. Realization of a dynamic model. Since different degree of familiarity with the word refers to different value of memory attenuation coefficient β, this algorithm constructs a functional relation between the degree of familiarity (FV) and memory attenuation coefficient.

It is assumed that FV have no extreme value and the β-FV function is convergent to β= 0.4307 (FV → +inf) and β= 0.1056 (FV → -inf). In addition, when FV=100 (original value), the corresponding β= 0.2038 (in accordance with the cogitive level of 'ambiguous'). An arc-tangent function is appropriate to describe the relationship between FV and β:

$$\beta(i) = C_1 \cdot \arctan(C_2 \cdot FV(i) + C_3) + C_4 \tag{5}$$

where C_1, C_2, C_3, C_4 are constants.

The result of fitting is shown in the following Figure 2 (x axle refers to FV; y axle refers to β):

Therefore, the mathematical expression for the β-FV function could be approximately treated as:

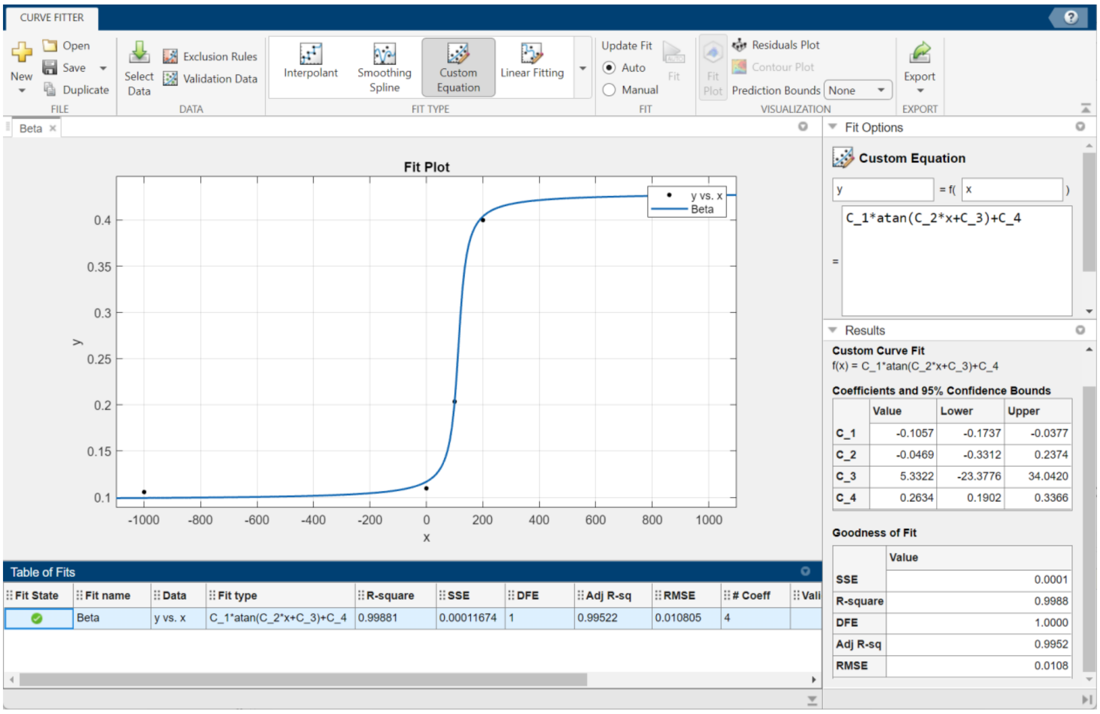

Figure 2. Result of Curve fitting for β-FV function (Picture credit: Original)

$$\beta(\mathrm{i}) = -0.106\arctan(-0.047FV(i) + 5.33) + 0.263 \tag{6}$$

2.3. Procedure of the program

2.3.1. Initialization of the program. At the very beginning of the process, the program is initialized where it reads the lexicon customized by users and inputs it into a table-type global variable, so that its elements could be handled as string or double type variables.

2.3.2. Main body of the program. According to the real-time value of FV, IMPT and M, the discrete probability density function is constructed which is used to determine the probability of occurrence of each word included in the lexicon. Next, a random word is decided complying with the distribution of the function where simultaneously the meaning of the word in Chinese and English, as well as its phonetic symbols, is exhibited to the learner. In addition, the Chinese meaning, the English meaning and the phonetic symbols are separately presented instead of being shown together. It is recommended that each time the learner reviews a word, he or she should only refer to one of them (CH meaning, EN meaning, phonetic transcription), in order that the retrieval strength could be lower, facilitating the formation of firmer memory.

Then, the learner enters the spelling of the word as the input. Subsequently the program receives the input and check the spelling and gives the response to the learner (Correct or not). If the answer is correct, the FV will diminish while if the answer is wrong, the FV will increase and the total error times of this word is recorded. After that, the probability density function will change according to the variance of FV and M before the next random word is given to the learner. The flow chart for the program is shown in the Figure 3.

2.3.3. Termination of the program. Since the program is the real-time hourly based system, the learner could decide to end learning at any time. When the CLOSE button is activated, real-time values are written into the excel file. Figure 3 showed the flow chart of main structure.

DOI: 10.54254/2753-8818/31/20241089

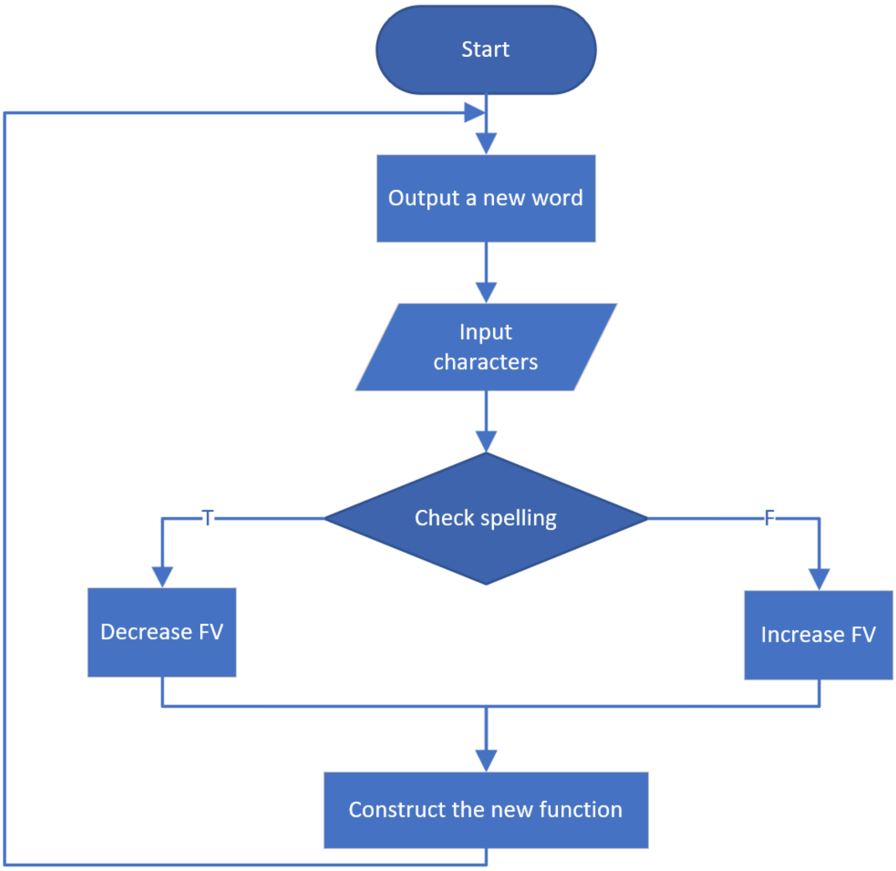

Figure 3. Flow chart of main structure (Picture credit: Original)

3. Results

3.1. The initial interface of the app

The initial interface of the app is shown below in Figure 4.

Function of interactive blocks:

'CLOSE' button: to write the ad hoc struct-type variable into the excel file and exit.

'Opt' button group: to decide the form of content shown in 'Explanation' block where if 'CHS' is selected, the 'Explanation' block will show the meaning of the certain word in simplified Chinese, similarly, 'EN' in English, and 'PT' shows the phonetic transcription of that word.

'Explanation' text area: to show different content according to 'Opt' button group.

'Note' text area: to show note bound with the word uploaded by the user. This text area is editable.

'On-Off' switch: to enable or disable 'Note' function.

'UPDATE' button: to upload the edited content in 'Note' text area to renew the note.

'Enter text' text area: where the user enters the spelling of the word.

'CHECK' button: to check whether the spelling of the word is correct or not.

'Result' text area: to show the result of the spelling.

DOI: 10.54254/2753-8818/31/20241089

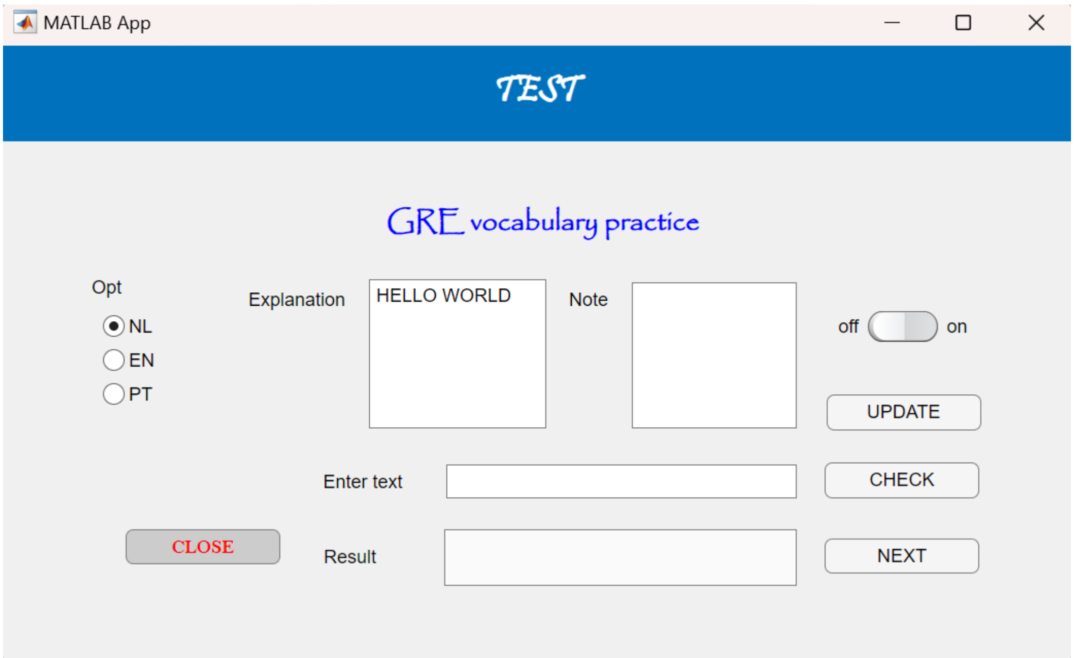

Figure 4. Initial interface of the application (Picture credit: Original)

3.2. Results of the program

Press 'NEXT' button to transfer to a new random word included in the lexicon.

It is assumed that the present word is 'amenity' and the results of the program is shown in the Figure 5 and Figure 6:

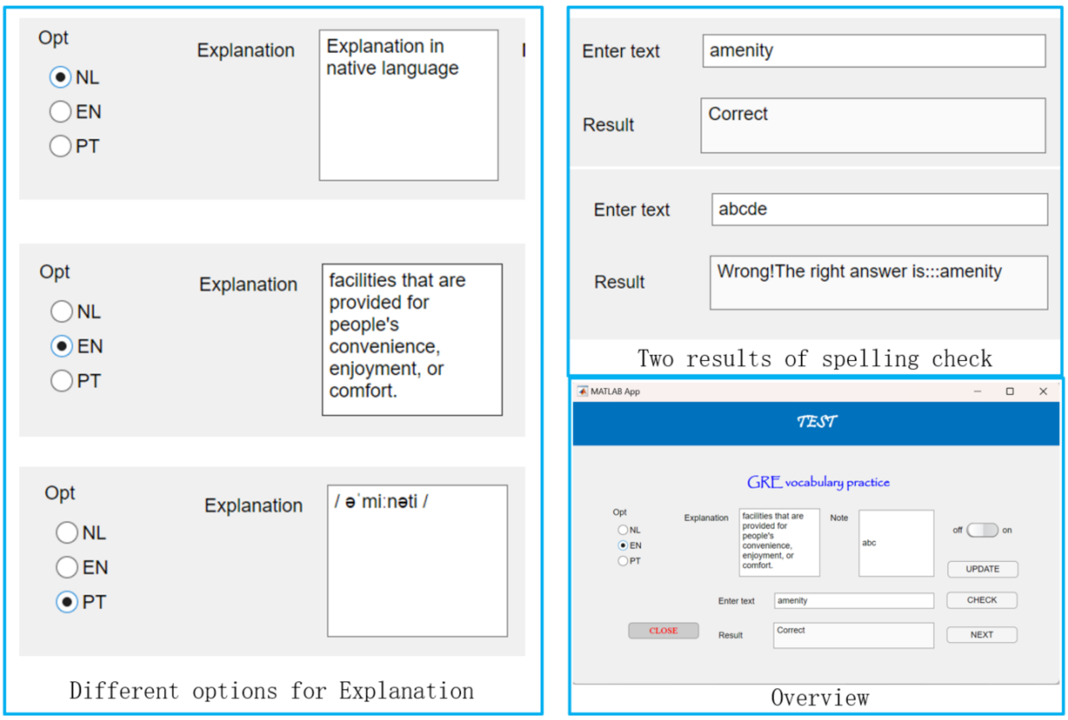

Figure 5. The results of the program interface (Picture credit: Original)

DOI: 10.54254/2753-8818/31/20241089

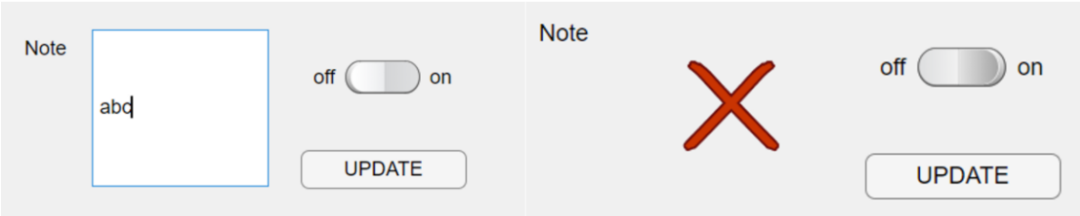

Figure 6. Demonstration for Note function (Picture credit: Original)

3.3. *Management of the lexicon*

No extra management system is implemented in this app except an excel file responsible for all date storage. The user could directly insert or delete a word in the lexicon and set corresponding value of the word either manually or automatically. (If a blank is left, the program will automatically give the default value.) The exemplar of lexicon management in excel is shown below in table 1.

Table 1. Exemplar of the lexicon management system stored in Excel

No.	Word	Explanation	Explanation /Other	Phonetic Transcription	Note	IMPT	FV	Last Revision
1	apostate	someone who has abandoned their religious faith, political loyalties, or principles	Explanation in native language	/ əˈpɒsteɪt /	a	3	80	2023.10.27.19
2	affidavit	written declaration made under oath	Explanation in native language	/ ˌæfəˈdeɪvɪt /	b	2	80	2023.10.27.19
3	amenity	facilities that are provided for people's convenience, enjoyment, or comfort.	Explanation in native language	/ əˈmiːnəti /	abc	3	100	2023.10.29.19
4	atone	make amends for(vt.); do penitence(vi.)	Explanation in native language	/ əˈtəʊn /	d	3	80	2023.10.29.16
5	impuissance	powerlessness revealed by an inability to act	Explanation in native language	/ ɪmˈpjuːɪsəns /	/	3	40	2023.10.29.16

3.4. *Background data visualization*

The probability density function of occurrence of each word is shown below in Figure 7 (taking the data in table 1 as an example):

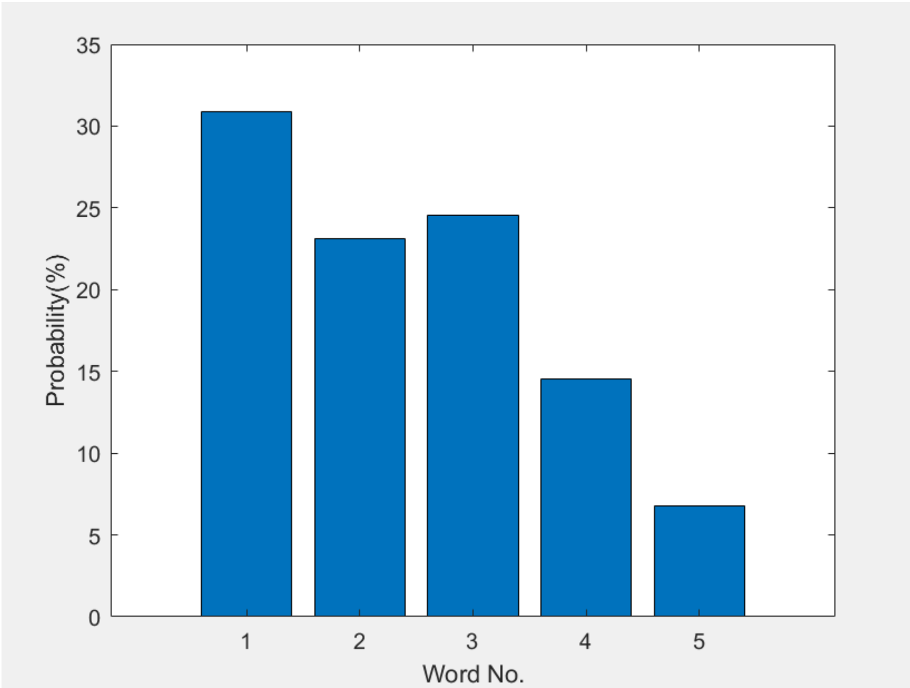

Figure 7. Visualization of discrete probability density function (Picture credit: Original)

Due to the synthetic influence of familiarity, priority and the forgetting curve, the probability of occurrence ranks in the sequence of: Word (1) > Word (3) > Word (2) > Word (4) > Word (5)

4. Discussion

As is shown in Figure 8, the probability density of occurrence for each word follows the pattern concerning with familiarity, priority and time interval of revision. Word (1) and word (2) share the same time interval of revision and FV, while word (1) is set to be more important than word (2), so the probability of its occurrence is higher than that of word (2). Similarly, word (2) and word (4) have identical FV and priority, but different time interval, so that P (2) is higher than P (4). Word (5) has the lowest probability to occur due to a low FV (high cognitive level).

Prospective works: 1) Consideration of mathematical model of spaced repetition: The weight coefficients could be added to the original discrete probability density function which could be modified to suit the optimum revision time of each word. With the advance of spaced repetition model, the best time to review a certain word could be calculated through a trainable model [12]. And a function relation between the weight coefficient and the time interval of retrieval could be established to better suit the memory pattern of the user. 2) Construction of vocabulary network: The construction of vocabulary network is proven effective when vocabulary is acquired. As Mei (2005) puts in his thesis, 'theoretically speaking, lexical network should be able to improve L2 vocabulary learning outcome' [13]. Therefore, contriving a new function to correlate some synonyms or derivative words and build more profound knowledge network will be beneficial to vocabulary learning.

5. Conclusion

Today, a bunch of English vocabulary learning applications could be found in current app market, most of which have handicaps in the vocabulary review section, including but not limited to: high retrieval strength of certain review methodologies which is detrimental to the formation of a firm long-term memory, and an absence of freedom to arrange flexible review tasks.

In order to resolve those problems, this paper presents a new method to review learned vocabulary through the orthographical practice, where a probability density function of occurrence is constructed

DOI: 10.54254/2753-8818/31/20241089

for each word. This method provides with an hourly-based dynamic revision system in order to enable the users to achieve learning and stop learning at any random time, as well as remove some inefficient method of reviewing vocabulary. However, verifying the practical effectiveness and quantifying the amelioration requiring a long span of time and a large-scale experiment involving numerous respondents.

References

[1] Wilkins D A 1972 *Linguistics in Language Teaching* (Massachusetts: MIT Press)

[2] Beijing Afan Technology Co., LTD 2016 China post-00s Internet Learning Behaviour Report. R/OL.

[3] Peng W 2021 Application of Ebbinghaus Forgetting Curve in English Vocabulary Teaching in Secondary vocational schools. *Proc. Symp. on innovation in Science and Education* (01), 184-5

[4] Wang G 2011 A study of conceptual and operational models of second language vocabulary knowledge *China. J. Foreign Language Education* (01), 51-6

[5] Dunlosky J, Rawson K A, Marsh E J, Nathan M J and Willingham D T 2013 Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. *J. Psychol. Sci. Public Interest* 14(1), 4-58

[6] Roediger H L and Karpicke J D 2006 Test-enhanced learning: taking memory tests improves long-term retention. *J. Psychol. Sci.* 17(3), 249–55

[7] Rowland C A 2014 The effect of testing versus restudy on retention: a meta-analytic review of the testing effect. *J. Psychol. Bull.* 140(6), 1432–63

[8] Bjork R A and Kroll J F 2015 Desirable difficulties in vocabulary learning. *Am. J. Psychol.* 128(2), 241–52

[9] Ebbinghaus H 1913 *Memory: A contribution to experimental psychology* (New York: Teachers College Press)

[10] Cai L, Xiong W and Sun X 2016 Adaptive vocabulary memorization model based on forgetting curve. *J. Microcomputer Applications* (05), 16-9

[11] Donkin C and Nosofsky R M 2012 A power-law model of psychological memory strength in short- and long-term recognition. *J. Psychological science* 23(6), 625–34

[12] Settles B and Meeder B 2016 A trainable spaced repetition model for language learning. *Proceedings of the 54th annual meeting of the association for computational linguistics* (01), 1848-58

[13] Mei X 2007 Association and second language vocabulary acquisition. D. (Shanghai: Shanghai International Studies University)