Evaluation and Development of Onomatopoeia CAPTCHAs

Michihiro Yamada
Riko Shigeno
Hiroaki Kikuchi
Graduate School of Advances Mathematical Sciences
Meiji University
Tokyo, Japan
cs172001@meiiji.ac.jp

Maki Sakamoto
Graduate School of Informatics and Engineering
The University of Electro-Communications
Tokyo, Japan

***Index Terms*—CAPTCHA, Onomatopoeia, Manga**

I. Introduction

The Completely Automated Public Turing Test to Tell Computers and Humans Apart (CAPTCHA) is a type of challenge–response test used for avoiding malicious software during automated registration. It plays an important role in security because fraud with computer agents is serious. The requirements of CAPTCHAs are:

1) a human can easily solve a CAPTCHA,
2) a computer cannot solve a CAPTCHA, and
3) CAPTCHAs can be generated automatically.

As deep learning technology has been developed, many of the existing CAPTCHAs were compromised and fail to satisfy condition (2).

In this study, we propose a new "onomatopoeia CAPTCHA" that applies *onomatopoeia*; i.e., words containing sounds similar to the noises they describe. Humans usually understand onomatopoeia unconsciously and use it in daily conversation; thus, it is clearly easy for humans to solve. However, it is difficult for computers because the mechanisms to recognize onomatopoeia are not very clear even now [1].

One of the difficulties of CAPTCHA schemes is the lack of reliable accuracy metrics. Some of the existing works deal with successful rate defined as a fraction of correctly answered tests. However, if we modify schemes as more complicated so that condition 2) is satisfied, then it may be hard to be solved by human resulting failure of condition 1) . So, we need to balance the tradeoff of two conditions.

To address the above issues of CAPTCHA scheme, we introduce two evaluation metrics, Human Acceptance Rate (*HAR*) and Machine Acceptance Rate (*MAR*), measured through comprehensive experiments. To balance both acceptance rates, we try to improve *HAR* with the five proposed schemes looking for the best scheme that allows humans solve CAPTCHA easily. Similarly, we attempt to reduce *MAR* as smaller as possible, that is, to make CAPTCHA unbreakable against attackers. Our experiment is evaluated by 63 Japanese and 63 foreigners participating from 16 countries. One of the proposed style of CAPTCHA is based on the Manga comics with onomatopoeia that may be recognized wide range of subject without suffering form language barrier and hence it helps to extend the coverage of users.

Our contribution of this work is as follows.

- A new CAPTCHA scheme using Onomatopoeia that is able to be synthesized from system.
- A comprehensive evaluation of accuracy metrics with respects to both human and machine (*HAR* and *MAR*), with five styles of queries and five smart attackers.
- An evaluation made be by a broad domain of subjects with distinct background knowledge in the world including 63 Japanese and 63 non-Japanese.

II. Experiment

A. Purposes of experiments

The purposes of these experiments are as follows.

1) Experiment 1: To clarify the difference between human and machine acceptance rates according to the question style.
2) Experiment 2: To compare the accuracy of the proposed onomatopoeia CAPTCHA with the existing CAPTCHA.
3) Experiment 3: To compare the accuracy of the onomatopoeia CAPTCHA with the "Manga onomatopoeia CAPTCHA".

B. Query styles

In this paper, we study several query styles for onomatopoeia CAPTCHA to find the best style in terms of *HAR* and *MAR*.

Style 1: (One-out-of-four-choices) Choose the appropriate onomatopoeia to express the given meaning. Example: Choose the onomatopoeia that means "walking". 1. Paku-paku, 2. Boso-boso, 3. Toko-toko, 4. Gya-ha-ha. (Answer: 3. Toko-toko.)

Style 2: (Onomatopoeia-vs-onomatopoeias) Choose the best appropriate onomatopoeia to express similarly the given onomatopoeia. Example: Choose the onomatopoeia that is similar

978-1-5386-7493-2/18/$31.00 © 2018 IEEE

TABLE I
EXPERIMENTAL RESULTS OF ACCEPTANCE RATIOS FOR STYLES 1–4

Styles	1			2		3	4	
Type of onomatopoeia	Dictionary	SK	SKALL	Dictionary	SK	Dictionary	Dictionary	SK
Subjects	Japanese			Japanese		Japanese	Japanese	
Number of subjects N	63			63		63	63	
Number of queries M	18	3	3	18	3	12	24	3
MAR_b	1/4	1/4	1/4	1/4	1/4	1/16	1/4	1/4
MAR_g	4/9	1/3	1/4	3/7	1/4	1/16	-	-
MAR_{gs}	0.400	0.333	0	0.277	0.333	-	-	-
MAR_d	1	1/4	0	1	1/4	1	1	1/4
Mean *HAR*	0.891	0.757	0.656	0.705	0.444	0.680	0.894	0.513
Standard deviation	0.084	0.229	0.290	0.107	0.259	0.162	0.063	0.280
Mean answering time [sec]	00:07	00:08	00:08	00:08	00:09	00:07	00:07	00:09

to "Guzu-Guzu"? 1. Hoku-Hoku, 2. Hera-Hera, 3. Horo-Horo, 4. Tara-Tara. (Answer: 4. Tara-Tara.)

Style 3: (Radio-button) Check whether the given meaning matches the onomatopoeia for each candidate. Example: Check if the given meaning "Fun" matches to the onomatopoeia for the following candidates. 1. Baku-Baku, 2. Uki-Uki, 3. Gonyo-Gonyo, 4. Ran-Ran. (answers: 2. Uki-Uki; 4. Ran-Ran.)

Style 4: (Choice-of-images) Choose the onomatopoeia that best represents the atmosphere of the given image.

Style 5: (Choice-of-manga) We hide the onomatopoeia used in manga comics and have subjects choose the one that would best match the hidden (red) portion. We denote by 5J and 5E if style 5 uses manga in Japanese or English, respectively. Example: which word is likely to fit in the red painted image? (Fig. 1)

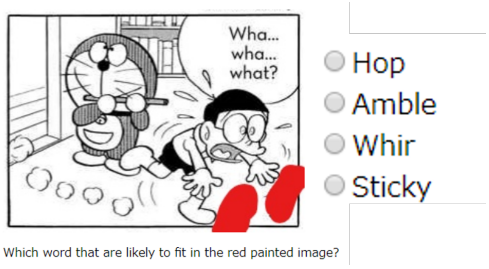

Fig. 1. Example of style 5E (Answer: Hop) [2]

C. Experimental method

We conducted two experiments. In the first experiment, we tested a total of 63 subjects including 54 people recruited via a crowdsourcing site and nine students from our laboratory (Experiments 1 and 2). Subjects were asked to answer a total of 84 questions on styles 1 through 4 at our web site for the experiments, and we stored the answers and the time required to complete. In the second experiment, 57 Japanese and 63 non-Japanese subjects (Experiment 3) participated. We asked the Japanese people 10 questions about 5J and we asked the non-Japanese 10 questions about 5E.

TABLE II
EXPERIMENTAL RESULTS FOR ACCEPTANCE RATIOS IN STYLE 5

Language	J	E_j	E_e
Subjects	Japanese	Japanese	non-Japanese
Number of subjects *N*	57	57	63
Number of queries *M*	10	10	10
MAR_b	1/4	1/4	1/4
MAR_g	-	-	-
MAR_{gs}	-	-	-
Mean *HAR*	0.895	0.452	0.410
Mean answering time[sec]	00:14	00:26	00:18

D. Experimental results

The experimental results for styles 1 through 4 are shown in Table I. The experimental results for style 5 are shown in Table II. The I table shows that style 4 has the highest *HAR* of almost 0.9. Style 1 has *HAR* of more than 0.8 while styles 2 and 3 have *HAR* of 0.6. We found no significant difference in the average answer time for the different styles. When synthetic onomatopoeias are used in styles 1, 2, and 4, the average *HARs* are on average 0.25 lower. Among the results for styles 1–4 (Table I) and style 5J (Table II), the *HAR* of Japanese manga onomatopoeia CAPTCHA of style 5J was the highest (*HAR* 0.895). From the experimental results of $5E_j$ and $5E_e$, the *HAR* of English onomatopoeia CAPTCHA was as low as that for Japanese manga onomatopoeia CAPTCHA (*HAR* 0.452 and 0.410).

REFERENCES

[1] Y.Shimizu, R.Doizaki, T.Kagitani and M.Sakamoto, "A system to Generate Onomatopoeia Corresponding to User's Impressions", JSAI Journal, Vol 30, No. 1, pp.319–330, 2015 (in Japanese).
[2] Fujio.F.Fujiko (2013), "DORaEMON selection2", Japan: Shogakukan, p.8.