Effects of Game on User Engagement with Spoken Dialogue System

Hayato Kobayashi **Kaori Tanio** **Manabu Sassano**

Yahoo Japan Corporation

9-7-1 Akasaka, Minato-ku, Tokyo 107-6211, Japan

{hakobaya, katanio, msassano}@yahoo-corp.jp

Abstract

In this study, we examine the effects of using a game for encouraging the use of a spoken dialogue system. As a case study, we developed a word-chain game, called *Shiritori* in Japanese, and released the game as a module in a Japanese Android/iOS app, *Onsei-Assist*, which is a Siri-like personal assistant based on a spoken dialogue technology. We analyzed the log after the release and confirmed that the game can increase the number of user utterances. Furthermore, we discovered a positive side effect, in which users who have played the game tend to begin using non-game modules. This suggests that just adding a game module to the system can improve user engagement with an assistant agent.

1 Introduction

Making users actively utter queries is important in a spoken dialogue system since they are generally not familiar with speaking to a system compared to typing on a keyboard. There have been several studies based on *gamification* for addressing this problem (Jurgens and Navigli, 2014; Gustafson et al., 2004; Hjalmarsson et al., 2007; Bell et al., 2005; Rayner et al., 2010; Rayner et al., 2012). Gamification is a concept of applying game design thinking to non-game applications, leveraging people's natural desires for socializing, learning, mastery, competition, achievement, and so on. However, it takes much time and effort to gamify a whole system, i.e., to consider how to design a game-like framework and combine new and current systems.

We therefore explore the possibilities of using of a game instead of gamifying a whole system. In other words, we address the question of whether a small module of an existing dialogue game can make users actively use the whole system. To this end, we developed a word-chain game as a case study and released the game as a module in the running Android/iOS app *Onsei-Assist* (Yahoo! JAPAN, 2015), which we describe later. We analyzed the log of user utterances after its release and confirmed that our results clearly answer this question positively.

The following are our main contributions.

- We analyzed vast amounts of dialogue data, i.e., more than tens of millions of user utterances cumulated via a running app of a spoken dialogue system.

- We discovered that just adding an existing game module to the system can have a positive impact on the non-game modules of the system from a case study of a word-chain game. This suggests that a game can help increase user engagement with an assistant agent.

The remainder of this paper is structured as follows. In Section 2, we introduce related studies on gamification for natural language processing systems. In Section 3, we briefly describe a spoken dialogue app, Onsei-Assist, whose log was used throughout our analysis. In Section 4, we explain how we developed a word-chain game module using a crowdsourcing service and in Section 5, we analyze the effects of using the game in Onsei-Assist. We conclude the paper in Section 6.

2 Related Work

We now briefly describe related studies on gamification for natural language processing systems, especially for spoken dialogue systems. When a gamified system is completely a game, the system is called a *game with a purpose (GWAP)*, or a *serious game*. Although a GWAP is sometimes differ-

entiated from gamification, we do not differentiate them for simplicity.

There have been many studies involving gamification for annotation tasks including anaphora resolution (Hladká et al., 2009; Poesio et al., 2013), paraphrasing (Chklovski and Gil, 2005), term associations (Artignan et al., 2009), and disambiguation (Seemakurty et al., 2010; Venhuizen et al., 2013). Recent studies (Vannella et al., 2014; Jurgens and Navigli, 2014) showed that designing linguistic annotation tasks as video games can produce high-quality annotations compared to text-based tasks.

There are several GWAPs based on spoken dialogue systems. DEAL is a game with a spoken language interface designed for second language learners (Hjalmarsson et al., 2007). In the NICE fairy-tale game system (Gustafson et al., 2004), users can interact with various animated characters in a 3D world. This game yields a spontaneous child-computer dialogue corpus in Swedish (Bell et al., 2005). CALL-SLT is an open-source speech-based translation game designed for learning and improving fluency, which supports French, English, Japanese, German, Greek, and Swedish (Rayner et al., 2010; Rayner et al., 2012).

However, each of these games or gamified systems was custom-made for a certain purpose, and to the best of our knowledge, we are the first to examine the effects of an existing dialogue game with an entertainment purpose, i.e., word-chain game, to a non-game system, especially in a spoken dialogue system.

3 Onsei-Assist

We used the log of a Japanese Android/iOS app of a spoken dialogue system, Onsei-Assist (Yahoo! JAPAN, 2015), throughout this analysis. Onsei-Assist is a Siri-like personal assistant developed by Yahoo Japan Corporation, where "Onsei" means "voice" in Japanese. It produced more than 20 million of utterances within a year of release on April 2012 via pre-installs to smartphones and downloads (more than one million) in GooglePlay.

Onsei-Assist was developed based on a client-server architecture, where the main system consists of four servers: a speech recognition server, meaning understanding server with natural language processing, response generation server, and speech synthesis server. The processing flow is as follows. A client, or smartphone, sends voice sig-

nals from a microphone to the speech recognition server and receives a recognition result of the user utterance in textual form; consequently, it sends the text to the meaning understanding server. This server differentiates the meaning of the utterance from the text and extracts information of variables such as named entities (e.g., Tokyo) and numerical expressions (e.g., 2014). It then generates a response pattern and sends it to the response generation server, which completes a response text by obtaining the required information via the API of several services. It also returns the response text with its prosody calculated by the speech synthesis server.

Onsei-Assist supports more than 20 services, each of which are launched by triggers based on natural sentences such as

- Route search ("From Osaka to Tokyo", "When does this train arrive?"),

- Weather information ("Today's weather", "Will it rain tomorrow?"),

- News ("News about the general election"),

- Web/image search ("Search for Tokyo Tower"),

and so on. In addition to such task-oriented dialogue modules, it can have a chat with users for general queries such as "How old are you?" and "Hello". Our system generates a response by choosing one from a pre-defined sentences based on a rule-based algorithm and learned model.

Table 1 shows examples of the log of user utterances, each of which is a tuple of five elements, i.e., (Time Stamp, User ID, Type, User Utterance, System Response). We obtained the log of more than 13 million utterances of 489 thousand users for our analysis.

4 Word-chain Game

First, we explain a word-chain game called *Shiritori* in Japanese. The principle of the word-chain game is to say a word based on rotation so that its head character is the same as the tail character of the previous word, e.g., (appl**e**, **e**e**l**, **l**i**p**, **p**ine, ...). It is a well known speech dialogue game in Japan since a syllable is basically represented by a character of a Japanese syllabary, i.e., Hiragana. The concrete rule used in this analysis is that each player must say a word based on rotation satisfying the following four conditions:

Time Stamp	User ID	Type	User Utterance	System Response
2014-8-1 12:34	UserA	Weather	今日の天気 (Today's weather)	晴れです (It's sunny)
2014-8-2 22:22	UserB	Chat	こんにちは (Hello)	こんにちわ (Hello)
2014-8-3 23:00	UserC	Chat	ありがとう (Thank you)	いえいえ (Your welcome)

Table 1: Example of log of user utterances.

1. The head of the word must be the same as the tail of the previous word.

2. The word must be a noun.

3. The word must not be a word already said in the game.

4. The tail of the word must not end with "ん (n)".

Conditions 2 and 3 prevent the game from being too long, and condition 4 is set because there is no word whose head character is "ん (n)" in Japanese.

Next, we explain the development of a word-chain game module for Onsei-Assist. We used a crowdsourcing service for obtaining words that people would usually use in the game because we worried that unfamiliar words extracted from Wikipedia and dictionaries could seriously deteriorate user satisfaction from a practical standpoint.

The process of collecting words is as follows. We prepared 1,150 seed words from dozen of employees in our company by using a simple word-chain game program developed only for this purpose. We created a crowdsourcing task asking workers to answer an appropriate word for each seed word based on the above rule. We repeated the task three times. Table 2 lists the results of the task for each repeated stage. Since the crowdsourcing service we used does not allow us to add a rule-check mechanism, we checked whether the results followed the rule after the task finished. About 90% of the answers were correct. Finally we obtained a sufficient amount of words (6,148) with their frequencies. We extracted the top 20 words based on frequency for each of the 66 Japanese head characters in the extracted words. This prevented the game from being too difficult since the workers rarely answered with words whose tail character was rare in Japanese. For example, the dictionary has only two words for the character "ぴ (pi)". Therefore, users can easily win by aiming for such a tail character.

We developed a word-chain game module for Onsei-Assist using the above dictionary. Figure 1

Stage	#Words	#Answers	#Errors
1	1,403	3,379	71
2	2,951	9,314	826
3	6,148	25,645	2,285

Table 2: Results of crowdsourcing task for obtaining possible words obeying word-chain game rule. #Words, #Answers, and #Errors represent number of distinct words, workers' answers, and errors due to breaking of rules, respectively.
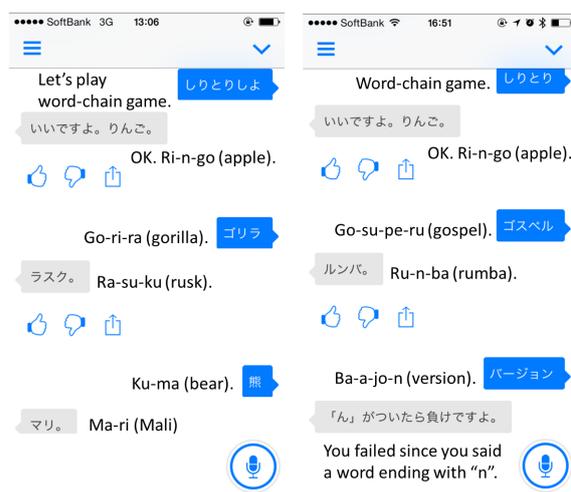

Figure 1: Screen shots when playing word-chain game module. Right and left balloons in each screen shot represent user and system's utterances, respectively.

shows two screen shots when playing the word-chain game module. In the module, the game starts by a user's trigger utterance such as "しりとり (Word-chain game)". The system replies with a response such as "いいですよ。りんご (OK. Ri-n-go)", and a user needs to say a word whose head character is "ご (go)" as the response. If the user says something that does not follow the rule, the system replies an error message such as "しりとりになっていません。(It's not a chained word)". The user can stop the game by using an ending word such as "ギブアップ (Give up)".
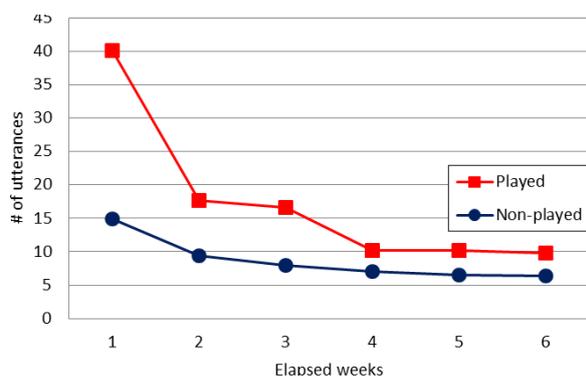
Figure 2: Average number of utterances over new users versus elapsed weeks. `Played` and `Non-played` represent users who had played and had not played the game on the first day, respectively.

5 Log Analysis

We conducted an analysis based on short- and long-term effects. For short-term effects, we define the reply rate of a system response R as the rate of the number of replies, which were uttered in a short period by users who received R, per the number of times R occurs in the log. The period was set to 20 minutes. We obtained a reply rate of more than 90% for every system response in the word-chain game. This is quite high, considering the fact that even a question-type system response "どうしました？ (What's happening?)" is about 80%. This implies that the game leverages users' natural desires for competition. In fact, the reply rates after a user won or failed (especially for saying a word already said) were 90.22% and 95.78%, respectively. This clearly indicates that users tend to retry to win after they failed.

For long-term effects, we averaged the number of utterances in a week over new users. Then we plotted it against elapsed weeks as shown in Figure 2, where `Played` and `Non-played` represent users who had played and had not played the game on the first day, respectively. We regard users who have not used the system over the last two months as new users to obtain sufficient data. The table clearly indicates that `Played` tended to use the system more frequently than `Non-played`. We also examined the difference between before and after game plays of active users. Table 3 shows the average number of utterances over game plays of active users in the week before and after each game. For extracting active users and obtaining a fair evaluation, we

	Before	After
(a) # of game plays	29,448	
(b) # of utterances	724,416	1,491,125
(c) # of game utterances	0	206,940
$((b) - (c))/(a)$	24.60	43.61

Table 3: Average number of utterances over game plays of active users week before and after each game play.

only considered game plays such that a user corresponding to each game play had used the system at least once and had not played the game for a week before game play. We found that game plays increased the average number of utterances by about 150% (from 24.60 to 43.61) despite the fact that we excluded utterances about the game. Note that these results are basically better than the results on new users in Figure 2 since we focused on active users. A possible reason is that users have become more familiar with this assistant agent through playing the game. Thus they began to use non-game modules more frequently.

6 Conclusion

We examined the effects of using a game for encouraging the use of a spoken dialogue system. We developed a word-chain game, called *Shiritori* in Japanese, as a case study and released the game as a module in a running Android/iOS app, Onsei-Assist, based on a spoken dialogue technology. We analyzed the log after the release and confirmed that the game can increase the number of user utterances. Furthermore, we discovered a positive side effect, in which users who have played the game tend to begin using non-game modules. This implies that a game can help to improve user engagement with an assistant agent. In other words, it is important to consider adding an entertaining module, such as a game, when developing a spoken dialogue system, as well as a useful module such as a route search.

For future research, we will examine other games such as a word association and quiz games. Since a game can be regarded as a simplification of a complex mechanism for natural dialogues, we hope to obtain generalized knowledge for improving spoken dialogue systems, if we can clarify which game can effectively improve which module in such systems.

References

Guillaume Artignan, Mountaz Hascoët, and Mathieu Lafourcade. 2009. Multiscale Visual Analysis of Lexical Networks. In *Proceedings of the 13th International Conference on Information Visualisation (IV 2009)*, pages 685–690. IEEE Computer Society.

Linda Bell, Johan Boye, Joakim Gustafson, Mattias Heldner, Anders Lindström, and Mats Wirén. 2005. The swedish NICE corpus - spoken dialogues between children and embodied characters in a computer game scenario. In *INTERSPEECH 2005 - Eurospeech, 9th European Conference on Speech Communication and Technology*, pages 2765–2768. ISCA.

Tim Chklovski and Yolanda Gil. 2005. Improving the design of intelligent acquisition interfaces for collecting world knowledge from web contributors. In *Proceedings of the International Conference on Knowledge Capture*, pages 35–42. ACM.

Joakim Gustafson, Linda Bell, Johan Boye, Anders Lindström, and Mats Wirén. 2004. The NICE Fairy-tale Game System. In *Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue (SIGDIAL 2004)*, pages 23–26. Association for Computational Linguistics.

Anna Hjalmarsson, Preben Wik, and Jenny Brusk. 2007. Dealing with DEAL: a dialogue system for conversation training. In *Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue (SIGDIAL 2007)*, pages 132–135. Association for Computational Linguistics.

Barbora Hladká, Jiří Mírovský, and Pavel Schlesinger. 2009. Play the Language: Play Coreference. In *Proceedings of the Joint Conference of the Association for Computational Linguistics and International Joint Conference of the Asian Federation of Natural Language Processing (ACL-IJCNLP 2009)*, pages 209–212. Association for Computational Linguistics.

David Jurgens and Roberto Navigli. 2014. It's All Fun and Games until Someone Annotates: Video Games with a Purpose for Linguistic Annotation. *Transactions of the Association of Computational Linguistics*, 2(1):449–464.

Massimo Poesio, Jon Chamberlain, Udo Kruschwitz, Livio Robaldo, and Luca Ducceschi. 2013. Phrase detectives: Utilizing collective intelligence for internetscale language resource creation. *ACM Transactions on Interactive Intelligent Systems*, 3(1):3:1–3:44.

Emmanuel Rayner, Pierrette Bouillon, Nikolaos Tsourakis, Johanna Gerlach, Yukie Nakao, and Claudia Baur. 2010. A Multilingual CALL Game Based on Speech Translation. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2010)*. European Language Resources Association (ELRA).

Manny Rayner, Pierrette Bouillon, and Johanna Gerlach. 2012. Evaluating Appropriateness Of System Responses In A Spoken CALL Game. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012)*, pages 2690–2694. European Language Resources Association (ELRA).

Nitin Seemakurty, Jonathan Chu, Luis von Ahn, and Anthony Tomasic. 2010. Word Sense Disambiguation via Human Computation. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, pages 60–63. ACM.

Daniele Vannella, David Jurgens, Daniele Scarfini, Domenico Toscani, and Roberto Navigli. 2014. Validating and Extending Semantic Knowledge Bases using Video Games with a Purpose. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL 2014)*, pages 1294–1304. Association for Computational Linguistics.

Noortje Venhuizen, Valerio Basile, Kilian Evang, and Johan Bos. 2013. Gamification for Word Sense Labeling. In *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013)*, pages 397–403. Association for Computational Linguistics.

Yahoo! JAPAN. 2015. Onsei-Assist (in Japanese). http://v-assist.yahoo.co.jp/.