

Prototyping Radiobots – – Automatic Radio Talks Generator Considering Live Feedback from Listeners

Rafal Rzepka¹, Yasutomo Kimura², Yoshihito Tsuji² and Keiichi Takamaru³

¹ Hokkaido University, Japan
rzepka@ist.hokudai.ac.jp

² Otaru University of Commerce, Japan
{kimura,ytsuji}@res.otaru-uc.ac.jp

³ Utsunomiya Kyowa University, Japan
takamaru@kyowa-u.ac.jp

Abstract. This paper presents a prototype of the Radiobot system which is characterized by artificial radio personalities able to automatically generate a vocal conversation and alter its content depending on comments from listeners. Preliminary experiments show that our proposed interaction model has promising features when compared to popular one-to-one assistant agents and conventional radio broadcasting.

1 Introduction

Non-task oriented dialog systems have been always a niche in comparison with task-oriented ones. However, except entertainment values [11], they can be an effective source for collecting conversations for further processing. They also can become a useful mean for supporting different purposes like second language learning [5] or cognitive therapies [8]. Popular voice assistants on various devices as Siri⁴, Google Now⁵, Cortana⁶ or Alexa⁷ are gradually being equipped with more and more chat-like responses, but various companies try to develop chatbots with purposes rather than particular goal oriented assistants with chatting features. Tay⁸, a bot developed by Microsoft Research for learning interactions, is one of such examples. However, as many inappropriate utterances learned from users showed, losing control over a bot may lead to serious problems. Our idea is to take users' input into consideration, but not to allow listeners to take control over the agent's conversation which is based on the latest news. We created a system, where one agent reads a piece of news from a news site and another

⁴ <http://www.apple.com/ios/siri/>

⁵ <https://www.google.com/search/about/learn-more/now/>

⁶ <https://support.microsoft.com/en-us/help/17214>

⁷ <https://www.amazon.com/Amazon-Echo-Bluetooth-Speaker-with-WiFi-Alexa/dp/B00X4WHP5E>

⁸ <http://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction>

agent uses fillers or comments on it, and sometimes react to listeners feedback received during the program via commenting function (radio program is streamed through an online broadcasting site). This allows us to experiment with a new type of radio broadcasting where main conversation is automatically generated and performed by artificial hosts and human users are (to some extent) allowed to influence the contents of the chatbots' dialog.

1.1 State of the Art

We believe that mixing informative and entertaining aspects should increase usefulness of interactive applications of this kind. Lack of content in chatbots was problematic since ELIZA [12], which was programmed to keep its user talking without providing any concrete and useful information. It has been shown that even simple commonsensical statements are preferred than such a type of utterances [9], however, even if bots utilizing Web resources can generate natural and on-topic output [2], it is rather difficult to call it informative or useful. [3] proposed a system which deals with this problem and generates not only longer but also amusing and semantically appropriate monologues which are Twitter-based. When conversational multiagents are considered, several approaches exist and work on virtual museum guides [6] is one of many examples. [10] created a system where two robotic agents act as standup-comedians but while museum guides rely on manually created scenarios, the comedians concentrate on their own dialog not accepting any input from the audience. The only algorithm for generating radio programs we managed to find was proposed by [4], but it does not involve any talking agents, it just learns user's music preferences. There was an idea [1] to make a system that actively changes a movie scenario according to user's input but was not realized because of high costs and insufficient technological resources (currently this approach is utilized in video games).

1.2 Aims of the Project

Our Radiobots project aims at a) mechanizing radio program contents generation, b) creating AI radio personalities and c) accepting user input which can influence the radio program scenario. To the authors' best knowledge, such system has never been created. The system is also supposed to become a proving ground for our parallel research projects including information retrieval, affective processing, morality assessment, humor generation and understanding, fun / topic attractiveness level measurement, poetry generation or cognitive biases discovery. We are also interested in solving problems with possible malicious input from anonymous users by using automatic moral consequences estimation [7]. Choosing an interesting comment worth reacting is not an easy task and requires involvement of several algorithms. As the first step we created a proof of concepts with a small number of functions and asked users to evaluate it by comparing it to ordinary radio programs and popular automatic assistants used in smartphones.

2 System Overview

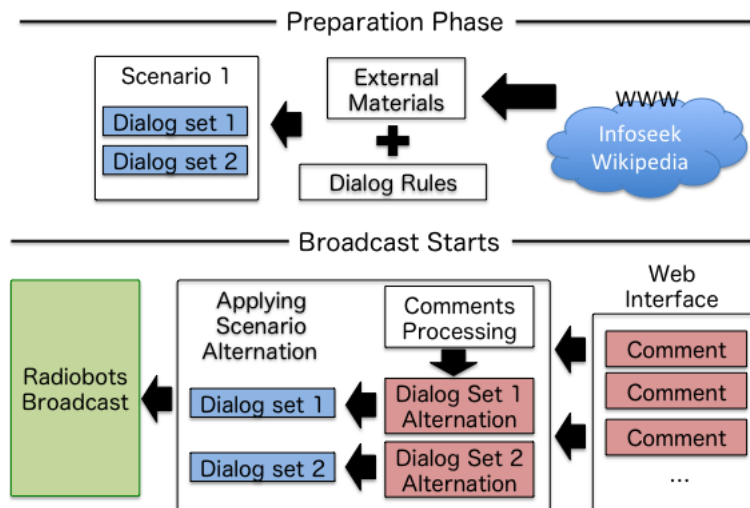


Fig. 1: Basic outline of the Radiobots system.

As explained above, to investigate users' impressions of the proposed new type of interaction, we built a prototype of our system. The idea is to develop an automatic dialog generator for multiple (two-three) agents which are hosts on an online radio program and discuss various topics. During the broadcast listeners are allowed to send comments instead of making phone calls or sending postcards to a radio station. As human DJs can react to e.g. tweets from listeners in real time, our artificial personalities should be able to react and even alter their conversations (or topic choices) accordingly if they decide to⁹. We created two radio personalities, purple-haired Otoha and red-haired Tae (see Figure 2), for which we used commercially available voices¹⁰. The prototype first retrieves current news articles from Infoseek¹¹ (by default the site displays 20 of them), then chooses two of them randomly, searches for terms / names in Wikipedia and extracts articles' first sentences (see Figure 1). Example using an article about investing trends in real estates, is given below¹²). These sentences are kept for later and used if a explanation is needed.

⁹ Theoretically automatic personalities can almost instantly analyze hundreds of comments and calculate e.g. trends which would be difficult for human DJs.

¹⁰ <http://www.ah-soft.com/voiceroid>

¹¹ <http://www.infoseek.co.jp>

¹² Radiobots work with Japanese language only, so the Japanese Wikipedia is utilized.

- 0:01: (Otoha) *Let's talk about "luxury condominium sales in the Tokyo metropolitan area is booming".*
- ...
- 0:24: (Otoha) *Is not an exaggeration to say that REIT is now underpinning the real estate market .*
- 0:33: (Tae) *Really?*
- 0:35: (Otoha) *So REIT... when big developers... while raising buildings...*
- 0:49: (Tae) *I couldn't find anything about developers on Wikipedia...*
- 0:57: (Otoha) *On the other hand, REIT, in fact, ... on public funds.*
- 1:08: (Tae) *A real estate investment trust (REIT) is a company that owns, and in most cases operates, income-producing real estate, isn't it?*

For the dialog generation we use ELIZA-based algorithm that is equipped with 857 utterances that always can be uttered (e.g. "interesting!", "nice!" or "I see"), 370 pronoun substitution rules ("you" \rightarrow "I", etc.) and 415 keyword-triggered reactions (e.g. "weather" \rightarrow "I wish it was always sunny"). Otoha (the main persona) reads a news article, and Tae (the commentator) uses above listed rules uttering follow-ups after one longer or two shorter sentences. Users who listen to the characters can send comments to the system using a nickname and an input text. From the set of comments, the program randomly chooses two-three comments and Tae reads them, while Otoha reacts with ELIZA-style utterances. At this moment the system is not equipped with comment choice capability and the user input is chosen randomly. For example when broadcast topic was "thallium sulfate", the interaction was as follows.

- Tae: *From user Kondoo we received a message "That's scary!"*
- Otoha: *It's not like that.*
- Tae: *From user John Hyakujiro we received a message "white powder..."*
- Otoha: *You can say that.*

3 Evaluation of Commentary Function

To evaluate the prototype we conducted a classroom experiment with 38 students (32 males and 6 females) who listened and interacted for two times (randomly chosen sets of news based dialogs were used) with Radiobots characters shown on a screen (see Figure 2). Every participant used two versions of the system - one with and one without commenting function.

In the second part of evaluation we decided to see how our system is regarded when compared to other types of active and not active conversations. Participants were asked to take another survey and compare the system with mobile assistants and conventional broadcasts on 7 points scale in four categories: tiresomeness, entertainment, understandability and usefulness; they were also asked to explain reasons for their choices.

The results are shown in Table 1. After confirming statistical significance with chi-squared (χ^2) test (see Table 2) and analyzing comments, we understood that when tiresomeness is considered, Radiobots bring less burden on users than

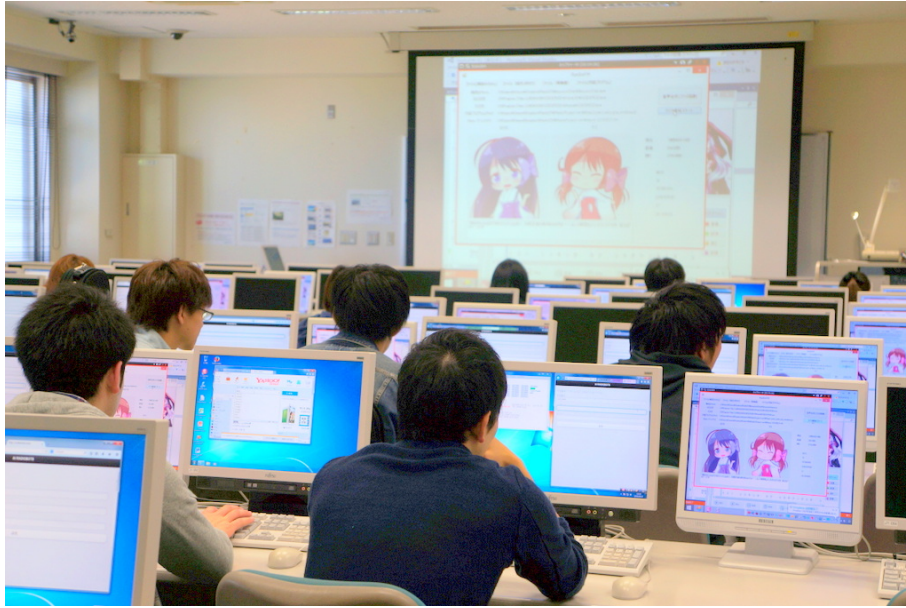


Fig. 2: Commentary function experiment.

mobile assistants but are just a bit more tiring than conventional radio broadcast scenario. As for entertainment values, possibility of commenting received more positive feedback than one input - one output style mobile assistants and traditional radio broadcasting.

The results showed that utilized voice generation, although using state-of the art software, were insufficient to be evaluated higher than their conventional counterparts. Some users commented that utterances are too flat and emotionless which suggests a need for more natural output. For many users usefulness of Radiobots was not clear, however they commented that it is an interesting new way for killing time.

4 Evaluation of Commentary Function

Because the commentary function is the main novelty of the proposed system, we also investigated how users felt about the possibility of commenting the broadcast content. We conducted an experiment, asking 172 students (115 males and 51 females,) to evaluate two systems - one with and one without commentary function. Randomly picked news categories for a version with commenting capability were “IT-related”, “Comic magazines” and “Celebrities”. For version without the function the utilized categories were “Business”, “Celebrities” and “IT-related”. A news set from 2015.12.10 was used and if the news entry was

Table 1: Seven points evaluation (-3 ~ +3): frequency distributions and averages.

Points assigned to RadioBots	when compared to 1-1 assistants				when compared to conventional radio			
	Tired.	Enter.	Under.	Usef.	Tired.	Enter.	Under.	Usef.
-3	0	1	2	3	2	1	2	1
-2	5	0	4	2	4	2	6	0
-1	1	1	6	3	12	2	9	5
0	8	2	6	6	10	2	9	7
+1	4	7	9	8	2	6	5	10
+2	12	15	8	8	2	16	4	9
+3	8	12	3	8	6	9	3	6
Total	38	38	38	38	38	38	38	38
Average	1.1	1.8	0.4	0.8	-0.1	1.5	-0.1	1.0

Table 2: Chi-squared (χ^2) test results.

7 points 3 groups	Differences with assistants				Differences with conventional radio			
	Tir.	Ent.	Und.	Use.	Tir.	Ent.	Und.	Use.
Negative (-3,-2,-1)	6 (0+5+1)	2 (1+0+1)	12 (2+4+6)	8 (3+2+3)	18 (2+4+12)	5 (1+2+2)	17 (2+6+9)	6 (1+0+5)
Neutral (0)	8	2	6	6	10	2	9	7
Positive (+1,+2,+3)	24 (4+12+8)	34 (7+15+12)	20 (9+8+3)	24 (8+8+8)	10 (2+2+6)	31 (6+16+9)	12 (5+4+3)	25 (10+9+6)
Total	38	38	38	38	38	38	38	38
χ^2	11.366 $p < .01$	33.964 $p < .01$	2.035 ns	7.929 $p < .05$	6.453 $p < .05$	23.280 $p < .01$	3.506 ns	11.612 $p < .01$

Table 3: Average values and t-test results (N=166).

Category	Without Comments	With Comments	t-test	Results
Entertainment	4.25	5.05	t(320)=5.25	p<.01
Understandability	3.67	3.66	t(320)=0.08	n.s.
Friendliness	3.88	4.52	t(320)=3.82	p<.01
Effortlessness	3.82	4.54	t(320)=4.40	p<.01
Agreeableness	3.33	3.45	t(320)=0.75	n.s.
Usefulness	4.41	3.77	t(320)=3.82	p<.01
Tiresomeness	3.41	3.36	t(320)=0.30	n.s.

lengthy, the system was set to cut it short to make it not longer than 90 seconds. After participating in 3 sessions for each version, the subjects completed a survey grading them on 7 points in following seven categories: entertainment, understandability, friendliness, effortlessness (easiness to deal with), agreeableness (degree of sympathizing), usefulness and tiresomeness. As shown in Table 3, in four categories the differences were significant. The system accepting users' comments appeared to be more interesting, friendlier and easier to sympathize with, but less useful when compared with the system without commentary function. Statistical significance was confirmed with IBM SPSS Statistics 21 software. Subjects were encouraged to freely write their opinions on the questionnaires. The differences in results are shown in Figures 3 and 4.

5 Conclusions and Future Work

In this paper we described a prototype of Radiobots, a non-task oriented dialog system that generates radio programs where agents chat with each other and that allows listeners to freely comment on these talks. After creating a prototype system we asked users to evaluate its performance by estimating levels of tiresomeness, entertainment, understandability and usefulness on seven points scale. The survey showed that in all four categories users prefer many-to-many interlocutors type system than a one-to-one type represented by mobile assistants. The proposed system scored higher than a human-generated radio programs in categories of entertainment and usefulness, and it got lower scores in tiresomeness and understandability. Because we focused on the commentary function, we have also experimentally confirmed the efficiency of this aspect. Although the Radiobots need improvement in speech generation, more sophisticated dialog processing and conversational strategies, and the experiments performed in a classroom are not perfect, the preliminary results suggest that the proposed system, as a hybrid of online broadcast and a chat system, is able to decrease user's burden of input while being sufficiently entertaining. Currently we work on system's sophistication to make it more attractive to users, also by adding modules for humor, poetry generation, metaphor processing, etc. This should help increasing interaction time and becoming more useful test environment for various AI experiments and for data collection.

References

1. Best, R.M.: Movies that talk back. *Consumer Electronics, IEEE Transactions on* 3, 670–675 (1980)
2. Higuchi, S., Rzepka, R., Araki, K.: A casual conversation system using modality and word associations retrieved from the web. In: *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*. pp. 382–390. Association for Computational Linguistics, Honolulu, Hawaii (October 2008), <http://www.aclweb.org/anthology/D08-1040>

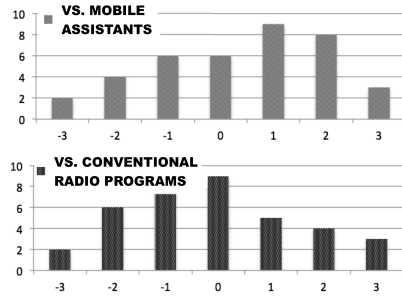


Fig. 3: Understandability level comparison (hard to understand \leftrightarrow easy to understand).

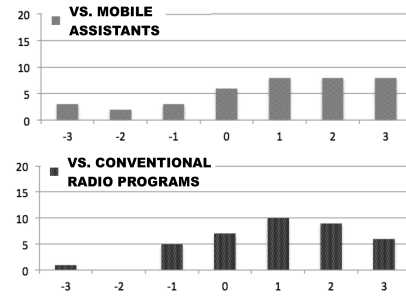


Fig. 4: Usefulness level comparison (useless \leftrightarrow very useful).

- Inaba, M., Yoshino, Y., Takahashi, K.: Open domain monologue generation for speaking-oriented dialogue systems (in japanese). Transactions of the Japanese Society for Artificial Intelligence 31(1), 1–9 (2016)
- Liu, N.H.: Design of an intelligent car radio and music player system. Multimedia tools and applications 72(2), 1341–1361 (2014)
- Mazur, M., Rzepka, R., Araki, K.: Proposal for a conversational english tutoring system that encourages user engagement. In: Proceedings of the 19th International Conference on Computers in Education. pp. 10–12. Asia-Pacific Society for Computers in Education (ICCE2011) (2012)
- Misu, T., Georgila, K., Leuski, A., Traum, D.: Reinforcement learning of question-answering dialogue policies for virtual museum guides. In: Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue, SIGDIAL '12. pp. 84–93 (2012)
- Rzepka, R., Araki, K.: Polarization of consequence expressions for an automatic ethical judgment based on moral stages theory. Tech. rep., IPSJ (2012)
- Rzepka, R., Araki, K.: ELIZA fifty years later: An automatic therapist using bottom-up and top-down approaches. In: van Rysewyk, S.P., Pontier, M. (eds.) Machine Medical Ethics, Intelligent Systems, Control and Automation: Science and Engineering, vol. 74, pp. 257–272. Springer International Publishing (2015), http://dx.doi.org/10.1007/978-3-319-08108-3_16
- Rzepka, R., Ge, Y., Araki, K.: Naturalness of an utterance based on the automatically retrieved commonsense. In: Proceedings of IJCAI 2005 - Nineteenth International Joint Conference on Artificial Intelligence, Edinburg, Scotland. pp. 996–998 (August 2005)
- Sjöbergh, J., Araki, K.: A complete and modestly funny system for generating and performing japanese stand-up comedy. In: Proceedings of COLING. pp. 111–114 (2008)
- Wallace, R.: Parsing the turing test, the anatomy of ALICE (2009)
- Weizenbaum, J.: Computer Power and Human Reason. W.H. Freeman and Co., New York (1976)