Comparing Modes of Information Presentation: Text versus ECA and Single versus Two ECAs*

Svetlana Stoyanchev[1], Paul Piwek[1], and Helmut Prendinger[2]

[1] NLG Group, Centre for Research in Computing
The Open University, Walton Hall, Milton Keynes MK7 6AA, UK
s.stoyanchev@open.ac.uk, p.piwek@open.ac.uk
[2] National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan
helmut@nii.ac.jp

Abstract. In this short paper, we evaluate the prospects of automatic dialogue script generation from text for presentation by a team of Embodied Conversational Agents (ECAs). We describe an experiment comparing user perception and preference between plain text and video ECA presentations modes and between monologue and dialogue presentation styles. Our results show that most users are not indifferent of the presentation mode and the user's preference is guided by the perceived understanding and enjoyment of the presentation.

1 Introduction

As devices for delivering information have become more and more powerful and portable (from SmartPhones and Tablet PCs to the iPad and Kindle), traditional paper-based solutions (including books, leaflets, newspapers, journals) for information presentation are gradually replaced by electronic delivery platforms. Electronic delivery of information opens up new opportunities for presenting information in ways that are more engaging for and adaptive to information consumers.

Traditionally, research in the area of Intelligent Information Presentation has focused on 'some level of internal representation' [13] from which the information is then presented in the most appropriate way. Much of the research on Embodied Conversational Agents (ECAs) for information presentation also relies on this assumption. In the field of Natural Language Generation (NLG), there has recently been a trend away from generation based on manually constructed inputs (usually in a knowledge representation language) to generation from widely available inputs. At least two complementary strands have emerged: data-to-text

* The research reported in this paper was carried out as part of the CODA (COherent Dialogue Automatically generated from text) project. The project was funded by the UK's Engineering and Physical Sciences Research Council under Grant EP/G020981/1. The authors would like to thank Ms Sara Winter for preparing the video materials for the experiment.

generation (D2T; e.g., [11]) and text-to-text generation (T2T/paraphrasing, e.g., [4, 12]).

In the current paper, we evaluate the prospects of the T2T approach to generate dialogue content for ECA teams. We investigate the efficacy of the resulting presentations by comparing original monologue text with generated dialogue script and also versions of the monologue and dialogue that are performed by one or two ECAs, respectively. Our aim is to compare the perceived quality of the different presentation modes and possible preferences by users for one mode or another.

The remainder of this paper is structured as follows. In the next section, we compare our approach with previous approaches to dialogue script generation and evaluation. Section 3 is at the heart of the current paper. Here we present the results from our user study, comparing monologue versus dialogue and text versus ECAs. Finally, Section 4 contains our conclusions.

2 Related Work

Several empirical studies show that delivering information in the form of a dialogue, as opposed to monologue, can be particularly effective for education [5] and persuasion [17]. However, none of these studies work with automatically generated dialogues. The current study is most similar to [19] where the authors compare presentation of generated information in monologue and dialogue audio mode. The authors generate presentation material from a relational database, in contrast to our approach where we generate dialogues from text.

Whereas most initial work on automatically generating dialogue scripts focused on input in the form of knowledge representations and the use of AI planning techniques [1, 18, 3], there has also been a parallel strand of research starting from text in monologue form, including Web2Talkshow [6] and the T2D system [8]. Our curent system is most similar to T2D: it also creates dialogue based on the intra and intersential discourse relations in the text, aiming to preserve the information of the input (rather than achieve a comic/humorous effect, as does Web2Talkshow). The main difference with T2D is that our system is based on discourse-to-dialogue mappings which are grounded in a parallel monologue/dialogue corpus. We have described the system [9, 14], the corpus [15] and extraction of the mappings [16] elsewhere and also performed a controlled study with four expert judges which showed that the automatically generated dialogue scripts (in text form) have both accuracy (i.e., whether the dialogue preserves the information from the input monologue) and fluency that is not worse than that of human-written dialogues [10]. Our current study aims to determine the potential of using ECAs to present automatically generated dialogues to users.

3 Evaluation and Comparison

We describe an online user study aiming to determine user preference between presentation modalities (video and text) and presentation styles (monologue

and dialogue). The experiment participants are presented with four presentation modes: text monologue, text dialogue, video of a single-character monologue, and video of a two-character dialogue.

3.1 Method

Materials The topic of the presentations is eco-driving. The presentation materials include an adopted version of *What is eco-driving?* [3] and *Game instructions* composed by the video game developers [7].[4]

Monologue	Dialogue
It's worth remembering modern cars are designed to set off straight away, so warming your engine is needless and wastes fuel. It also causes engine wear as does keeping the engine running when you're stationary. If you're stuck in traffic it's best to turn your engine off completely as most modern cars are designed to use virtually no extra fuel to re-start.	TEACHER: Warming your engine is needless and wastes fuel. STUDENT: Why is that? TEACHER: It's worth remembering modern cars are designed to set off straight away TEACHER: It also causes engine wear STUDENT: As does keeping the engine running when you 're stationary? TEACHER: Yes. STUDENT: What if you're stuck in traffic? TEACHER: It's best to turn your engine off completely

Table 1. Example of the presentation material.

We created four types of presentations for the materials: plain text monologue (original), text dialogue, single-character video monologue, and two-character video dialogue. To generate dialogues, we first manually parsed the discourse relations of the input monologues [2] and then used our M2D system [10] to generate dialogue. Since we are primarily interested in whether the sequence of dialogue acts proposed by M2D provides suitable content for ECA video presentations, we corrected manually any syntactic and semantic errors, whilst leaving the dialogue act sequence unchanged. Table 1 shows an example of a snippet from a monologue document and a corresponding dialogue that were presented to the experiment participants. The videos of ECA monologue and dialogue presentations were generated using *xtranormal* MovieMaker.[5]

Fig. 1. Agents in Video Dialogue Presentation.

[3] `http://www.guardian.co.uk/ford-econetic/driving-lessons`
[4] The participants did not play the actual game.
[5] `http://www.xtranormal.com/`

Non-verbal behaviours were handcrafted using the behaviour authoring tool provided by *xtranormal*. When creating videos, we used the same gestures for the speaking character in both conditions. Figure 1 shows an image from a dialogue presentation.

Questionnaire and Procedure The participants completed an on-line questionnaire on their personal systems. Each participant viewed four presentations, each in a different modality (video/text) and style (monologue/dialogue). Each presentation took about 2 to 3 minutes to view or read. After each presentation, the participants were asked to report whether they understood the presentation, found it engaging/enoyable/natural/fun. The ratings were made on a 5-point likert scale. The participants were prompted to make a choice for their preferences between each pair of the four presentation modes. The participants were requested not to interrupt while completing the questionnaire. After viewing all of the presentations, the participants were given an option to provide feedback about the presentations.[6]

Participants We recruited 40 volunteer participants using computational linguistics mailing lists, university mailing lists, and Facebook connections. The participants self-reported language skill as "I can fluently communicate in English".

Design We designed four groups to control for the presentation mode of a document (see Table 2). Each document was presented in different mode across the four conditions. Each participant viewed exactly one instance of each document in the same order (Doc1, Doc2, Doc3 and Doc4), each in a different presentation mode. The participants were distributed over the groups by adding each new participant to the group with fewest participants so far, resulting in 10 participants per group.

	Material	Group A	Group B	Group C	Group D
Doc 1	What is Eco-driving (part 1)	Video Dia.	Video Mono.	Text Dia.	Text Mono.
Doc 2	What is Eco-driving (part 2)	Video Mono.	Text Dia.	Text Mono.	Video Dia.
Doc 3	Game instructions (part 1)	Text Dia.	Text Mono.	Video Dia.	Video Mono.
Doc 4	Game instructions (part 2)	Text Mono.	Video Dia.	Video Mono.	Text Dia.

Table 2. Experiment design.

[6] The participants were also asked recall questions about the content of the presentations. We are currently analysing the recall results.

3.2 Results

Preference Table 3 shows the results for user preferences between the styles and the modalities. For video presentations, 40% of participants prefer dialogue, 45% prefer monologue, and 15% have no preference. For text presentations, the majority of the participants prefer monologue over dialogue (65% over 30%) and only 5% have no preference.

The tendency for the users in both monologue and dialogue modes is to prefer video, however for the dialogue presentations this tendency is stronger than for monologue with 62% of the users preferring video.

Compare Preference for Monologue/Dialogue Style				Compare Preference for Text/Video Modality			
modality	Prefer Dialogue	Prefer Monologue	No Pref	style	Prefer Video	Prefer Text	No Pref
Video	40.0%	45.0%	15.0%	Monologue	52.5%	42.5%	5.0%
Text *	30.0%	65.0%	5.0%	Dialogue *	62.5%	30.0%	7.5%

Table 3. Experiment results. * indicates statistically significant difference between the two preferences according χ^2 test ($p < .05$)

Compare Video Monologue and Dialogue Modes We compare the scores for the ratings between video monologue and dialogue presentations (see Table 4). The first row of the Table shows the scores for all of the participants. The second row show the scores for only those participants who indicated that they prefer video dialogue over monologue presentation. The third row shows the scores for those preferring video monologue over dialogue.

While the scores between presentation modes for all participants are not significantly different, the participants who preferred dialogue report better understanding and enjoyment for the dialogue than the monologue. On the other hand, those who prefer monologue, report better understanding and enjoyment for the monologue than the dialogue.

participants (number)	understand VD/VM	enjoy VD/VM	natural VD/VM	engaging VD/VM	fun VD/VM	see again VD/VM
All (40)	3.83/4.03	3.00/2.95	2.50/2.78	3.00/2.80	2.50/2.35	2.93/2.85
Prefer VD (16)	4.25/3.88**	3.50/2.94*	3.13/2.75	3.63/2.81**	3.13/2.44**	3.81/2.75**
Prefer VM (18)	3.56/4.28	2.67/3.11*	2.06/2.94**	2.67/2.89	2.11/2.44	2.22/3.06**

Table 4. Ratings for the video dialogue (VD) and video monologue (VM) presentations. Statistically significant difference tested with paired t-test is marked with * for $p < .05$ and ** for $p < .01$

3.3 Discussion

The results show that majority of the participants have a preference for mode of presentation. While for text presentation majority prefer monologue, for video presentation the participants are almost equally split between monologue and dialogue. Their preference correlates with self-reported understanding and enjoyment. The results suggest that some participants may prefer monologue because they find dialogue less natural. However, those who prefer dialogue, show no statistically significant difference in the naturalness scores between the dialogue and monologue styles suggesting that the preference is personal and subjective.

Presentation quality has a strong effect on the user choice between video and text modality. The participants commented on quality of video/audio as well as of text presentations. 45% of all participants found quality of audio in the presentations *bad* or *very bad*. This may have affected comprehension scores for video and led to a stronger tendency to prefer text over video. One of the participants wrote that *"Maybe if the quality of animation was more Hollywood-like, it would win over the written text."*

4 Conclusions

We compared user preference between two presentation modalities (monologue and dialogue) and two presentation styles (text and dialogue) and found that a majority of users have a preference for one of the two modalities and presentation styles. Additionally, a majority prefers to view dialogues as videos. For text presentations, monologue is preferred over dialogue. Finally, we found that the user's preference is correlated with their perceived understanding of the material and enjoyment of the presentation.

With improved quality of video and audio presentations, we expect that preferences will shift towards video. Since a clear majority of participants prefer a particular presentation mode, ideally an Intelligent Information Presentation system should allow people a choice between the monologue and dialogue modes. This in turn suggests a need for automatic generation of dialogue content for ECAs from text, with text still being one of the most ubiquitous information sources available.

References

1. E. André, T. Rist, S. van Mulken, M. Klesen, and S. Baldes. The automated design of believable dialogues for animated presentation teams. In *Embodied Conversational Agents*, pages 220–255. MIT Press, Cambridge, Mass., 2000.
2. L. Carlson and D. Marcu. Discourse tagging reference manual. Technical Report ISI-TR-545, ISI, September 2001.
3. M. Cavazza and F. Charles. Dialogue Generation in Character-based Interactive Storytelling. In *Proceedings of the AAAI First Annual Artificial Intelligence and Interactive Digital Entertainment Conference*, Marina Del Rey, California, USA, 2005.

4. T. Cohn and M. Lapata. Large margin synchronous generation and its application to sentence compression. In *Procs of EMNLP-CONLL*, pages 73–82, Prague, 2007.
5. S. Craig, B. Gholson, M. Ventura, A. Graesser, and the Tutoring Research Group. Overhearing dialogues and monologues in virtual tutoring sessions. *International Journal of Artificial Intelligence in Education*, 11:242–253, 2000.
6. A. Nadamoto and K. Tanaka. Complementing your TV-viewing by web content automatically-transformed into TV-program-type content. In *Proceedings 13th Annual ACM International Conference on Multimedia*, pages 41–50. ACM Press, 2005.
7. A. Nakasone et al. A 3D Internet based experimental framework for integrating traffic simulation and multi-user immersive driving. In *4th Int'l Conference on Simulation Tools and Techniques*, 2011.
8. P. Piwek, H. Hernault, H. Prendinger, and M. Ishizuka. T2D: Generating Dialogues between Virtual Agents Automatically from Text. In *Intelligent Virtual Agents*, LNAI 4722, pages 161–174. Springer Verlag, 2007.
9. P. Piwek and S. Stoyanchev. Generating Expository Dialogue from Monologue: Motivation, Corpus and Preliminary Rules. In *Proc. of NAACL*, Los Angeles, USA, 2010.
10. P. Piwek and S. Stoyanchev. Data-Oriented Monologue-to-Dialogue Generation. In *Proc. of ACL/HLT*, Portland, USA, 2011.
11. E. Reiter. An Architecture for Data to Text Systems. In *Procs of ENLG-2007*, pages 97–104, Schloss Dagstuhl, Germany, 2007.
12. V. Rus, A. Graesser, A. Stent, M. Walker, and M. White. Text-to-Text Generation. In R. Dale and M. White, editors, *Shared Tasks and Comparative Evaluation in Natural Language Generation: Workshop Report*, Arlington, Virginia, 2007.
13. O. Stock and M. Zancanaro, editors. *Multimodal Intelligent Information Presentation*, volume 27 of *Text, Speech and Language Technology*. Springer, 2005.
14. S. Stoyanchev and P. Piwek. The CODA System for Monologue-to-Dialogue Generation. In *Proc. of SIGDIAL*, Portland, USA, 2011.
15. S. Stoyanchev and P.Piwek. Constructing the CODA Corpus: a Parallel Corpus of Monologues and Expository Dialogues. In *Proc. of LREC*, Malta, 2010.
16. S. Stoyanchev and P.Piwek. Harvesting re-usable high-level rules for expository dialogue generation. In *Proc. of INLG*, Trim, Ireland, 2010.
17. S. V. Suzuki and S. Yamada. Persuasion through overheard communication by life-like agents. In *Procs of the 2004 IEEE/WIC/ACM International Conference on Intelligent Agent Technology*, Beijing, September 2004.
18. K. van Deemter, B. Krenn, P. Piwek, M. Klesen, M. Schroeder, and S. Baumann. Fully Generated Scripted Dialogue for Embodied Agents. *Artificial Intelligence Journal*, 172(10):1219–1244, 2008.
19. S. Williams, P. Piwek, and R. Power. Generating Monologue and Dialogue to Present Personalised Medical Information to Patients. In *Procs ENLG 2007*, pages 167–170, Schloss Dagstuhl, Germany, 2007.