

TABLEPRESENTER: A LIFE-LIKE AGENT EXPLAINING TABLES

Kazumasa KONDO
Graduate Program of Creation Informatics
Kyushu Institute of Technology
680-4 Kawazu, Iizuka, Fukuoka 820-8502, JAPAN
email: k_kondo@pluto.ai.kyutech.ac.jp

Kunio YAMAMOTO and Tsukasa NOMA
Department of Artificial Intelligence
Kyushu Institute of Technology
680-4 Kawazu, Iizuka, Fukuoka 820-8502, JAPAN
email: {kunio,noma}@ai.kyutech.ac.jp

ABSTRACT

This paper proposes a life-like presentation agent system called *TablePresenter*. In the *TablePresenter*, a user edits a table on Microsoft Excel, and also specifies the presentation contents for explaining the table in the form of tree structure, called *presentation tree*. The system then generates animated presentation by an agent. Presentation trees have intermediate nodes labeled as “contrast” and “compare,” and such types of nodes are associated with appropriate speech text templates with gesture specifications. Traversing the presentation tree then yields speech texts for explaining the table, and the resultant texts animate the agent. We developed a prototype *TablePresenter* system, and it successfully generated animated presentations for tables.

KEY WORDS

life-like agent, interface agent, pedagogical agent, virtual human, presentation, *TablePresenter*.

1 Introduction

Recently much attention has been paid to life-like interface agents[1]. Their use is, however, still limited to small number of applications/websites except for Microsoft Agents found on some Microsoft products[2]. This is because most of current agent systems are domain-specific ones with elaborate “intelligent” rules for, e.g. sightseeing guides and weather reports[3][4], or general-purpose systems with script-like languages[5][6][7].

Now let us suppose that a virtual agent would explain an object, e.g. a product, diagram, or table, instead of real humans. If the target object is different from that of existing domain-specific systems, we need to develop a brand-new intelligent system with huge costs, or to write an explanation script on a general-purpose system from scratch. This prevents us to extensively use life-like interface agents.

To solve the above-mentioned problem, this paper proposes a life-like agent system called *TablePresenter*, where all we have to do is to make a table on Microsoft Excel, and to diagrammatically specify presentation contents. Animated presentation for explaining the table by a life-like agent is then automatically generated from the above data. Tables are general-purpose information representations, and our system is thus achieves domain-

independence. At the same time, our diagrammatic presentation specification liberates us from tedious scripting work.

This paper is organized as follows: we first discuss the structure of tables and their explanations in Section 2. Section 3 overviews our system, and our prototype implementation is presented in Section 4. Experimental results are reported in Section 5, and finally Section 6 concludes this paper.

2 Tables and Their Explanations

A table is a systematic arrangement of data usually in rows and columns for ready reference[8]. Each item in the table is typically indexed by the topmost and/or leftmost item on the same row and/or column. For example, Table 1 shows a typical example representing monthly charges of ADSL (Asymmetric Digital Subscriber Line) communication line in Japan and Korea. In this example, the bottom-right item “4,076 yen” is indexed by “8M” and “Korea.”

We generally use such tables for showing tendencies, comparing items/rows/columns, and then make some conclusions. If a general-purpose scripting language is used to explain a table, we need to write a huge volume of scripts including four required data: (1) speech text, (2) reference items/positions on the table, (3) reference timings in the speech text, and (4) appropriate behaviors and gestures.

But if we use an information structure involved in a table, required specification for presentation will be greatly decreased. For example, in case of comparison, target items to be compared considerably determine their explanations. In Table 1, for example, if we choose “3,419 yen” and “3,377 yen” for comparison, it typically means that the comparison is made between 1.5 M line and 8 M line in terms of Japanese ADSL services. On the other hand, if we choose “3,419 yen” and “3,135 yen”, it means that the comparison is made between Japan and Korea in terms of

Country	1.5M	8M
Japan	3,419 yen	3,377 yen
Korea	3,135 yen	4,076 yen

Table 1. A sample table.

1.5 M service.

As shown in the next section, The *TablePresenter* effectively use such tabular structural information for generating animated presentation.

3 System Overview

In the *TablePresenter*, a table to be explained is first edited on a editor like a spreadsheet. Next, a *presentation tree*, which specifies presentation contents, is edited for the table. All the user has to manually make is only these two data. Figure 1 diagrammatically represents a sample presentation tree for Table 1.

In Figure 1, each rectangle, which is a leaf in the tree, represents an item on the table, and each oval, which is a non-leaf node in the tree, represents a “function” in the presentation. For example, the node “topic” represents the statement of a topic like a topic sentence. The node “contrast” represents contrasting its children nodes, and the node “compare” represents comparing its children nodes in terms of price, which is specified as domain. Dashed arrows between siblings represent the order of presentation.

For each type of non-leaf nodes, an appropriate speech text template with gesture specifications is associated with it. Sample associations are shown in Table 2, where gestures and their timings are omitted.

Traversing the tree with applying the templates yields a speech text for the agent to explain the table. Figure 2 illustrates how data items on the table and speech text templates are integrated in a bottom-up fashion to generate a speech text. In this case, the generated text is:

This table shows ADSL monthly charges. For 1.5 M, 3,419 yen of Japan is more expensive than 3,135 yen of Korea. But for 8 M, 3,377 yen of Japan is cheaper than 4,076 yen of Korea.

Underlines in the above text represent their pointing gestures by the agent. An animated presentation for the table is then generated from the speech text with gestures.

Among the four required data mentioned in Section 2, only (2) reference items/positions on the table are explicitly given in a presentation tree. The rest of three: (1) speech text, (3) reference timings in the speech text, and (4) appropriate behaviors and gestures, are all obtained through the integration of tabular information structures and the templates. This enables us to easily make animated presentations.

4 Implementation

To implement a prototype *TablePresenter* system, we used Microsoft Excel, VBS (Visual Basic Script), and XML. At first, a user edits a table on Excel, and simultaneously edits its presentation tree in XML (Figure 3). Then he/she runs a VBS program on the Excel that translates the presentation tree to its corresponding speech text with gestures.

```
<?xml version="1.0" ... ?>
<presentation-tree>
  <topic><title /></topic>
  <contrast>
    <compare domain="price">
      <cell>B2</cell>
      <cell>B3</cell>
    </compare>
    <compare domain="price">
      <cell>C2</cell>
      <cell>C3</cell>
    </compare>
  </contrast>
</presentation-tree>
```

Figure 3. A sample presentation tree in XML.

This table shows ADSL monthly charges. For 1.5 M, 3,419 yen of Japan is more expensive than 3,135 yen of Korea. But for 8 M, 3,377 yen of Japan is cheaper than 4,076 yen of Korea.

Figure 4. A generated speech text with gesture labels.

The translated speech text is shown in Figure 4. In this text, labeled phrases between <point> and </point> are positions where the corresponding items on the table are pointed by the agent when he/she speaks the phrases.

The Excel stores the table data as csv, from which an image of the table is generated. An agent makes presentations in front of the table image (Figure 5). Since the positions of items can be obtained in a graphical (world) coordinate system, the agent can point appropriate positions in the presentation.

In our implementation, we used Microsoft MSXML3.0 for processing XML, Microsoft Speech API 5.1 for voice output, OpenGL for graphics output, and an agent shape model from Viewpoint Corporation.

5 Results

We implemented a prototype system as mentioned above, and made some experiments on it. Animated presentations for tables were then successfully generated. Frames from the experiments are shown in Figure 5.

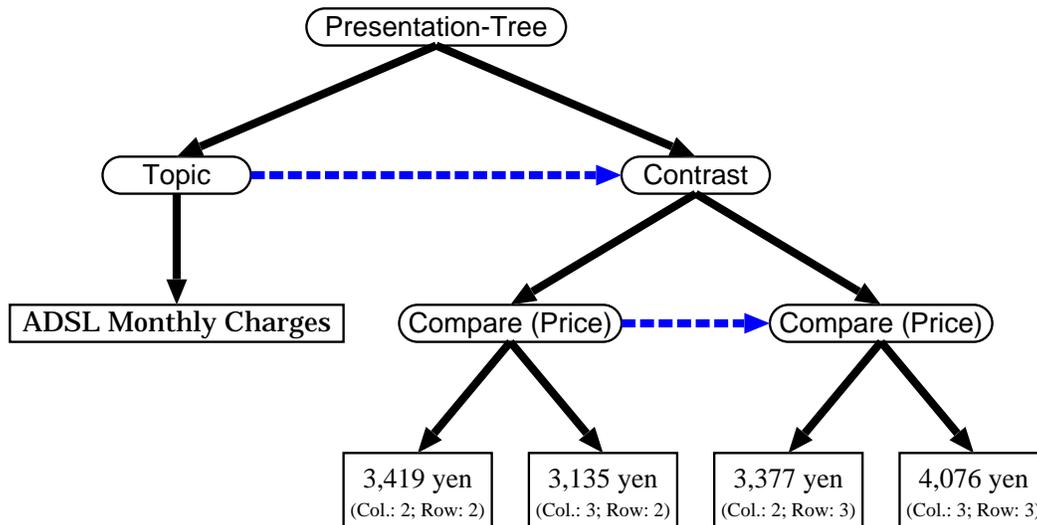


Figure 1. A sample presentation tree.

Node type	Speech text template
Topic	This table shows (CHILD).
Contrast	(CHILD1) But (CHILD2)
Compare (Price)	For (COMMONINDEX), (CHILD1) of (INDEX1) is {more expensive than / the same as / cheaper than } (CHILD2) of (INDEX2). (Note: phrases in { } are chosen by evaluating (CHILD1) and (CHILD2). (COMMONINDEX) is the index of the common column/row of (CHILD1) and (CHILD2). (INDEX1) and (INDEX2) are indexes of different column/row of (CHILD1) and (CHILD2), respectively. If both their rows and columns are different, another text template is used.)

Table 2. Node types and their associated speech text templates.

6 Conclusions

This paper proposed *TablePresenter*, where animated presentations for explaining a table are generated only from a simple presentation specification called *presentation tree*, which represents the *semantics* of the explanation.

To enable *TablePresenter* to generate animated presentations only from the presentation tree, we effectively used implicit information structures in a table. Although the target of the current system is limited to explanations of a table, our *TablePresenter* system is the first step toward life-like presentation agents achieving both domain-independence and laborsaving specification. Just as Cassell et al. combined speech texts to behavior expressions in [9], we combined presentation semantics of a table to its animated presentations.

On the other hand, we avoided *excessive* automation for generating presentations. For example, Mittal et al.[10] and Corio and Lapalme[11] would suggest fully automated presentation of tables/charts. But in reality, any tables/charts can be interpreted in various ways by various people, and we believe that at least the *semantics* of the presentation should explicitly be specified by the user.

Future work includes development of a presentation tree editor, and extension of target information structures. In our current prototype system, a presentation tree is manually edited on a XML editor. To really liberate users from tedious work of animated presentation generations, we need an environment where users can specify presentation trees in a graphical fashion, possibly integrated with table editors. For extension of target information structures, we should pay attention to charts/diagrams as well as more complex tables.

Acknowledgements

This work is partially supported by National Institute of Information and Communications Technology (NICT) through ASO Information System Co., Ltd., and by Japan Society for the Promotion of Science (JSPS) through Grant-in-Aid for Scientific Research (C) 17500072.

References

- [1] J. Cassell et al. (Eds.) *Embodied conversational agents* (Cambridge, MA: MIT Press, 2000).

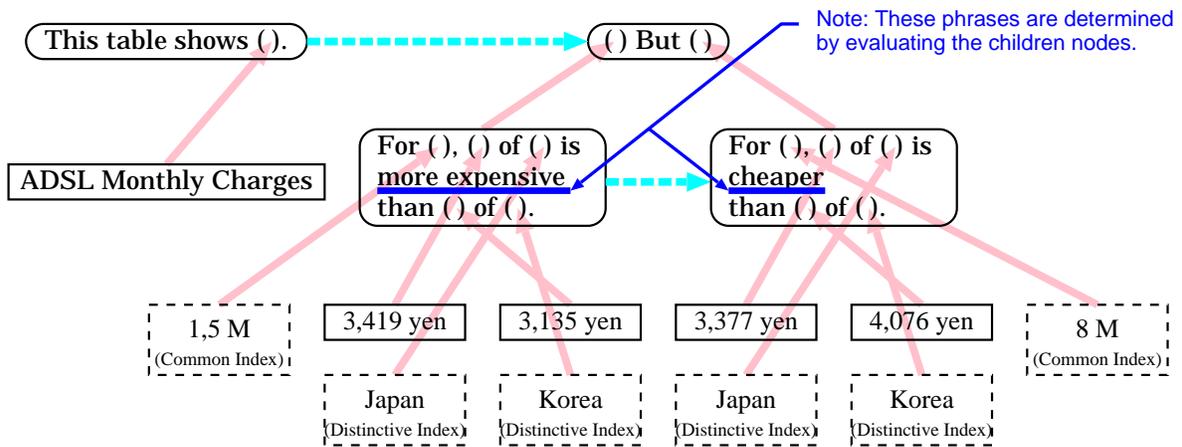


Figure 2. Presentation tree traversal and speech text generation. (Dashed rectangles are indexes on the same columns/rows.)

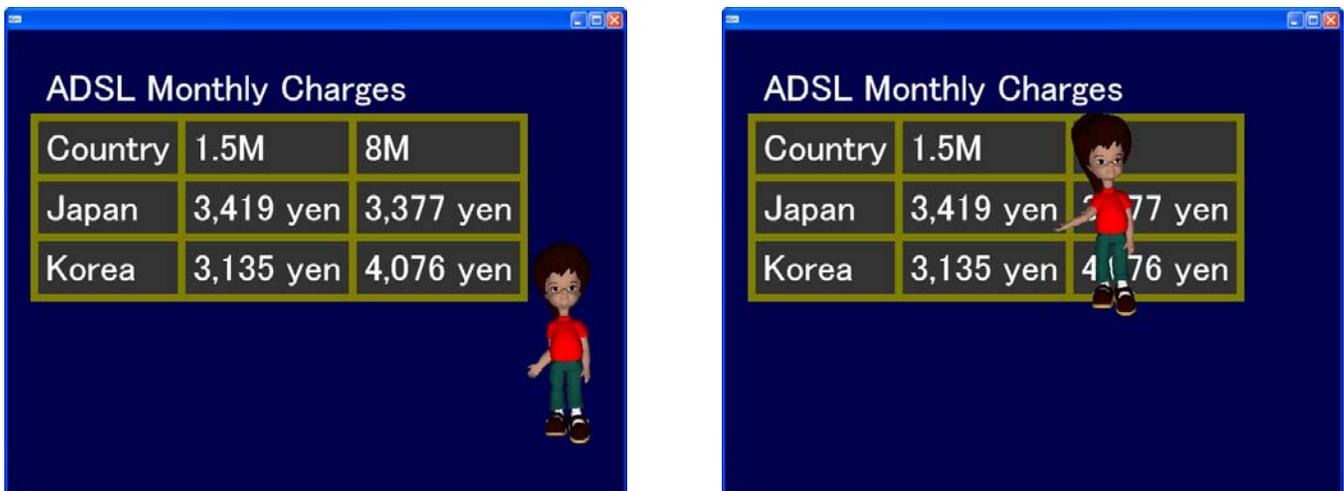


Figure 5. *TablePresenter* compares ADSL monthly charges between Japan and Korea.

- [2] Microsoft Corp., *Microsoft Agent home page*, <http://www.microsoft.com/msagent/>.
- [3] E. Andre, T. Rist, and J. Muller, Employing AI methods to control the behavior of animated interface agents, *Applied Artificial Intelligence*, 13(4), 1999, 415-448.
- [4] T. Noma, Y. Shiratori, and H. Baba, Automated generation of animated weather reports with a virtual weatherman, in S. Hashimoto (Ed.) *Multimedia Modeling (Proc. Int'l Conf. on Multimedia Modeling 2000)*, (Singapore: World Scientific, 2000) 413-427.
- [5] T. Noma and N.I. Badler, A virtual human presenter, *Proc. IJCAI-97 Workshop on Animated Interface Agents*, Nagoya, Japan, 1997, 45-51.
- [6] M. Hayashi, H. Ueda, and T. Kurihara, TVML(TV program Making Language) — Automatic TV Program Generation from Text-based Script —, *Imagina 99*, Monaco, 1999.
- [7] T. Noma, L. Zhao, and N.I. Badler, Design of a virtual human presenter, *IEEE Comput. Graph. Appl.*, 20(4), 2000, 79-85.
- [8] *Merriam-Webster Online Dictionary*, <http://www.m-w.com/>.
- [9] J. Cassell, H.H. Vilhjalmsson, and T. Bickmore, BEAT: the behavior expression animation toolkit, *Proc. SIGGRAPH 2001*, Los Angeles, CA, 477-486.
- [10] V.O. Mittal et al, Describing complex charts in natural language: A caption generation system, *Comput. Linguist.*, 24(3), 1998, 431-467.
- [11] M. Corio and G. Lapalme, Generation of texts for information graphics, *Proc. 7th European Workshop on Natural Language Generation*, Toulouse, France, 1999.