

Emotional Reading of Medical Texts Using Conversational Agents (Short Paper)

Gersende Georg

Centre des Cordeliers UMRS 872, Eq20
15 rue de l'Ecole de Médecine, F-75006

HAS, 2 Avenue du Stade de France
Saint-Denis La Plaine Cedex, F-93218
gersende.georg@spim.jussieu.fr

Catherine Pelachaud

University of Paris 8, INRIA
INRIA Rocquencourt, Mirages,
78153 Le Chesnay Cedex, France
catherine.pelachaud@inria.fr

Marc Cavazza

School of Computing
University of Teesside
TS1 3BA Middlesbrough
United Kingdom
m.o.cavazza@tees.ac.uk

ABSTRACT

In this paper, we present a prototype that helps visualizing the relative importance of sentences extracted from medical texts using Embodied Conversational Agents (ECA). We propose to map rhetorical structures automatically recognized in the documents onto a set of communicative acts controlling the expression of an ECA. As a consequence, the ECA will dramatize a sentence to reflect its perceived importance and rhetorical strength (advice, requirement, open proposal, etc). This prototype is constituted of three sub-systems: i) G-DEE, a text analysis module ii) a mapping module which converts rhetorical structures produced by the text analysis module into communicative functions driving the ECA animation and iii) an ECA system. By bringing the text to life, this system could help their authors (in our application, expert physicians) to reflect on the potential impact of the writing style they have adopted. The use of ECA re-introduces an affective element which cannot easily be captured by other methods for analyzing document's style.

Categories and Subject Descriptors

H.5.1 [Multimedia Information Systems] Animations; J.3 [Life and Medical Sciences]: Medical information systems; I.2.11 [Document and Text Processing]: Document Preparation - Markup languages - Hypertext/hypermedia.

General Terms

Algorithms, Human Factors.

Keywords

Embodied Conversational Agents, Document Engineering, Markup languages.

1. INTRODUCTION

The conversion of text to other modalities has been proposed initially as a means to facilitate access to its informational content. In recent years, the use of ECA to read aloud documents using Text-To-Speech (TTS) has gained increased popularity, due to progress in animation and speech synthesis. However, more sophisticated applications can be envisioned if one realises the potential of an ECA to reflect more than just the informational content of the text [2, 7, 13]. ECAs have been demonstrated to

Cite as: Emotional Reading of Medical Texts Using Conversational Agents (Short Paper), Georg G, Pelachaud C, Cavazza M. *Proc. of 7th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, Padgham, Parkes, Müller and Parsons (eds.), May, 12-16., 2008, Estoril, Portugal, pp. 1285-1288.

Copyright © 2008, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

bring added value (such as disambiguating text, adding communicative and affective information) to many applications for which a more human-like presentation [8] is beneficial, including assistance, help and guidance [1,2].

In this paper, we introduce a first prototype developed to visualize the importance of specific sentences within medical documents using an ECA. Clinical guidelines are normative texts, aimed at physicians, produced by various Health authorities, which promote best practice in Medicine, based on the concept of *evidence-based medicine*. They are complex documents which require significant amounts of specialized knowledge for their production. Clinical guidelines are based on the notion of *recommendation*, which are syntactic constructs associated to a strong rhetorical value. For instance, "The administration of low doses of aspirin (75 mg/day) is recommended for hypertensive patients with type 2 diabetes in primary care." One main challenge associated to the clinical guidelines' production is to be able to anticipate the impact of the specific recommendations they contain as a function of the style used. This is why we propose the automatic visualization of recommendations, as animating a recommendation through an ECA to restore the link between document content and the original committee discussion which decided on its formulation.

2. SYSTEM OVERVIEW AND ARCHITECTURE

The system presents itself as an ECA interface "reading aloud" specific recommendations extracted from a clinical guideline. It is actually constituted of three sub-systems: i) a document engineering environment, G-DEE [6] (Guidelines Document Engineering Environment) which automatically identifies the most relevant sentences of a guideline (the recommendations), ii) a mapping module which converts those recommendations into the communicative act format used by the ECA, a mark-up language known as APML [5] and iii) an ECA system called Greta [10]. The system operates as follows. Firstly, G-DEE is run offline to analyse the clinical guideline as a whole. It produces a document in which all recommendations are identified through a set of specific mark-ups for their operators and the contents they apply to (referred to as the *scopes* of the operator). An example of scopes marking-up is:

"<Front-scope> The administration of low doses of aspirin (75 mg/day) </Front-scope> <Op_Reco> is recommended </Op_Reco> <Back-scope> for hypertensive patients with diabetes type 2 in primary care </Back-scope>."

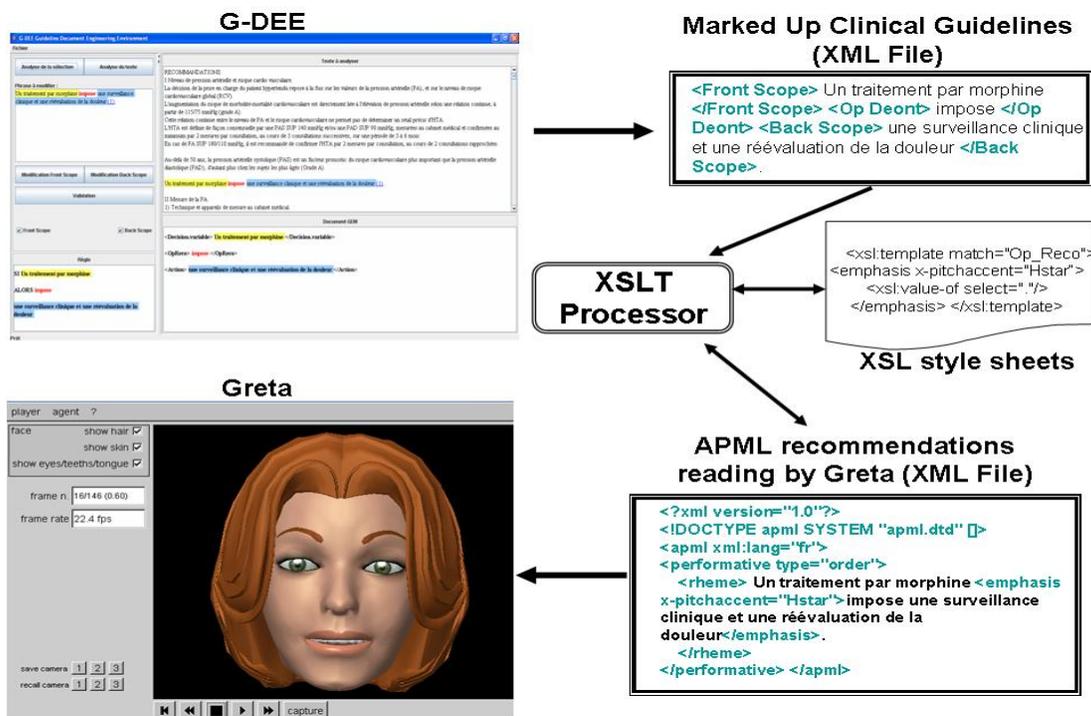


Figure 1. Overview of the architecture.

A marked-up recommendation appears as highlighted text in G-DEE (Figure 1). This text fragment can be selected interactively, which then triggers the generation of an APML file animating Greta on that sentence (the generation uses an XSLT conversion module). During this process, tags on communicative acts linked to recommendation strength are added to the text, as the result of the automatic mapping of rhetorical structures. Finally, Greta processes this APML file and utters the corresponding recommendation, displaying appropriate nonverbal behaviour, which reflects the importance of the recommendation and places emphasis on relevant scopes. In this way, the actual strength of the recommendation and its potential impact can be visualized.

3. AUTOMATIC EXTRACTION OF RECOMMENDATIONS FROM TEXTS

3.1 The Document Engineering Environment

We are interested in clinical guidelines that belong to the generic category of normative texts, to which much research has been dedicated. G-DEE [6] supports multiple document processing functions including the automatic recognition of recommendations using shallow NLP techniques recognizing *deontic operators* in medical texts such as “authorize”, “forbid”, “ought to”. Let us now consider the different aspects that determine the strength and emphasis of a recommendation. Firstly, deontic operators fall within the broad categories of permission, obligation or interdiction. Within these broad categories, specific deontic operators can be classified according to their “strength”. Strength is not just an issue of vocabulary, but relates also to syntactic constructs (which have been uncovered as part of the process of deontic operator extraction). In other words, that a specific drug “should not be used” is stronger than it being “not recommended”. It can also be noted that this concept bears

some similarity with the illocutionary strength of communicative acts (which constituted our initial inspiration for this project).

3.2 The Greta Platform

The Greta agent [10] used in these experiments is a platform developed for research in non-verbal behavior, including an animation system with facial parameters supporting detailed expressive animations synchronized to a TTS system. Greta’s animations are controlled using instructions in the APML language [4]. Communicative acts are gathered in classes depending on the information they convey [11]. In particular, previous study [12] have been conducted to elaborate the links between performatives (communicative acts), such as: suggest, propose, refuse, etc., and facial expressions. Three main classes of performatives have been considered: *request*, *inform* and *question*. Performatives of the class *request* have been characterized along three dimensions: i) to whom is the action requested, ii) how certain one is of the information being provided and iii) the power relationship between interactants [12]. Based on the representation of performatives along these three dimensions, we have proposed a mapping between each of these dimensions and the ECA’s facial expressions. That is, the facial expression associated to a given performative is obtained by combining the expressions arising from each dimension. Being certain or uncertain can be shown on the eyebrow region: one frowns when being very much certain of what one says, but raises eyebrows if uncertain. Head orientations (such as head kept straight up or tilted aside) can be a sign of a power relation: submissiveness is often shown by displaying our neck [4] while dominance is characterized by a straight up head). Performatives also contain an intrinsic emotional factor. A frown marks the performative ‘order’ as one can get angry if the interlocutor does not comply with the requested action.

4. RELATED WORK

Our work focuses on conversational agents for visualising rhetorical structures extracted from medical texts. It is related to storytelling agents [3], or emotionally expressive agents [1], with two important differences. The first one is that the ‘emotional’ content to be visualised is actually related to the importance and authority of a text fragment, rather than to its dramatic qualities. The second is naturally the application area, and the practical use of such a system to estimate the impact and readability of a given document style. It should be also emphasized that these documents have no less emotional impact because they’re directed at a physician’s audience: issues of importance, authority and responsibility generate powerful emotional responses as well. No work has yet reported the use of ECA to explore medical text perception by physicians. Clinical guidelines are based on the notion of *recommendation* which is a rhetoric structure advising or forbidding a specific course of action (from a pragmatic perspective this corresponds to a *deontic operator*). These recommendations have a significant emotional content which is linked to notions of authority and responsibility.

5. IDENTIFYING THE RHETORICAL STRENGTH OF RECOMMENDATIONS

Physicians do not always identify the most important information when they read clinical guidelines because of the variable quality of their formulation, and phenomena of ambiguity, imprecision, and vagueness [9]. The physician’s background has also been shown to play a role in their interpretation of guidelines [6]. In order to formalize the concept of rhetorical strength of a recommendation, we conducted a study involving 14 medical experts from INSERM (French National Institute for Health) and the French National Authority for Health (HAS). These experts rated the strength of 37 recommendations extracted from recent clinical guidelines published by the HAS. They ranked the strength of each recommendation according to a predefined six-point scale defined as follows:

- CAT1- well-identified best practice, which is compulsory
- CAT2- a practice well adapted to the clinical situation that presents demonstrable benefits
- CAT3- accepted practice which can be advised, or to be considered
- CAT4- a possible practice left to the discretion of the physician
- CAT5- a statement explaining a given clinical practice
- CAT6- a useful information item

Figure 2. Categories for evaluating the strength of recommendations.

For each deontic verb used in recommendations, we are able to associate a numerical score quantifying its rhetoric strength. This analysis will serve as a starting point to map the rhetorical strength of deontic expressions onto the emotional categories of Greta.

6. MAPPING RHETORICAL STRUCTURES ONTO MULTIMODAL COMMUNICATIVE ACTS

The process by which the rhetorical strength of textual recommendations will be visualized rests on a mapping from deontic operators onto multimodal communicative acts. These can be described as the dynamic expression of traditional speech acts

(order, advice, propose ...), using speech parameters and dynamic animation of non-verbal behavior, in particular facial expressions. The rationale for such a mapping derives from the pre-existing commonality between certain deontic operators, used in the description of recommendations, and the set of primitive speech acts embedded in the APML control language (which already contains speech acts such as *advise*), although the two were developed independently by different authors. This mapping attempts to generalize these commonalities by relating deontic operators to communicative acts, but also their perceived strength to the *rheme* part [14] of APML expressions, corresponding to the intentional structure that contains the new information. We have elaborated the mapping between the six categories of the strength scale and the performatives by considering the common values for these 3 dimensions (Figure 3).

- CAT1 (to impose / APML: ‘order’)** - only the frown is kept, as the other behaviours are also power signs. To highlight the importance, emphasis is added through head nods.
- CAT2 (to recommend / APML: ‘recommend’)** - represented by a less intense frown.
- CAT3 (to propose / APML: ‘advice’)** - displayed using the eyebrow shape (slight rising of the eyebrows).
- CAT4 (may / APML: ‘suggest’)** - characterized by raised eyebrows and tilted head.
- CAT5 (rarely indicate / APML: ‘inform + emphasis’)** - translated by looking at one’s addressee and performing a head nod on the emphasised word.
- CAT6 (should be suspected / APML: ‘inform’)** - displayed through gaze behaviour, namely looking at the addressee.

Figure 3. Mapping between strength and performative type.

The following example corresponds to Category 2. The dedicated style sheet enables to transform a marked-up recommendation to an APML format (Figure 4) that supports the mapping of the “*il est recommandé*” (“it is recommended”) deontic verb to the *recommend* performative type.

```
<apml> <performative type="recommend"> <rheme> <emphasis x-pitchaccent="Hstar">Il est recommandé</emphasis> de réaliser un écho-Doppler veineux lors de la prise en charge de tous les patients présentant un ulcère des membres inférieurs. </rheme></performative> </apml>
```

Figure 4. The resulting APML file corresponding to a recommend performative type.

The corresponding expression for Greta (Figure 5- left) consists of a recommendation with an emphasis on the deontic verb “*il est recommandé*” (it is recommended) and a raised eyebrow, while the *suggest* communicative act (Figure 5- right) is associated to a slight raising of the eyebrows and a head nod.



Figure 5. Expressions for recommend (left) and suggest (right).

7. PRELIMINARY USER EVALUATION

We conducted a preliminary evaluation of the system with 6 medical experts drawn from the group of the 14 experts that participated in the definition of recommendations' strengths. For this evaluation, 9 recommendations, automatically extracted by G-DEE, were visualized by Greta according to their rhetorical strength. The main objective of this evaluation consists of determining whether Greta improves the perception of the recommendations' strength, for instance by generating a stronger consensus or helping to disambiguate between categories. For this evaluation, we produced 9 videos representing Greta reading the 9 sentences with their corresponding communicative acts. These videos were presented to each of the 6 medical experts to rate the recommendation strength they perceived when Greta read the different recommendations. The average strength as well as the standard deviation were calculated for each recommendation, with and without Greta (Figure 6).

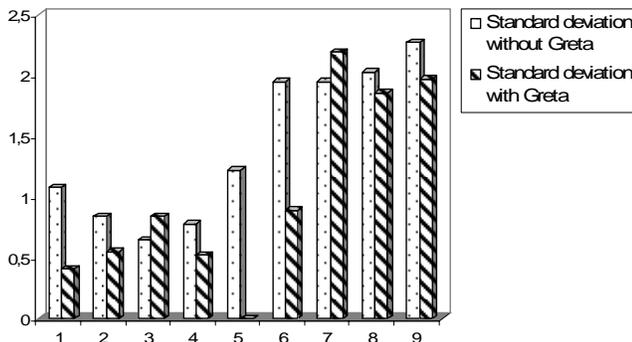


Figure 6. Impact of Greta on the standard deviation of experts' judgments of recommendations' strength.

8. DISCUSSION AND CONCLUSION

Finding the best formulation for a recommendation is a complex process, which often involves multiple cycles of discussion and negotiation within expert working groups. However, these revisions often take place after the initial document has been assembled. They are then disconnected from the consensus group discussions in which social and nonverbal behaviour plays an important part in highlighting the importance of specific recommendations. To a large extent, the system presented here can restore the link between the wording of a recommendation and its intended impact on the reader. It should help selecting the appropriate level of emphasis required, as well as balancing the importance of recommendations across the document as a whole. Our preliminary results suggest that Greta has an impact on the perception of recommendations strength. The significance of the overall distribution was tested by one-way ANOVA which showed this result to be statistically significant ($P < 0.0474$). Most importantly, we observed a significant effect of Greta on the standard deviation of perceived recommendations' strength, and that effect is more pronounced for intermediate categories, such as CAT3 and CAT5. We can argue that the diminution of the standard deviation with Greta corresponds to a better consensus between medical experts. These first results are encouraging and future work will consist of evaluating this approach with a larger test set of recommendations, also using more sophisticated expressive mechanisms such as gestures.

9. ACKNOWLEDGMENTS

Gersende Georg is partly funded through a post-doctoral fellowship from "Region Ile-de-France". We thank medical experts from the French National Health Authority (HAS) and INSERM (French National Institute of Health) for their participation in data collection and in preliminary evaluation experiments.

10. REFERENCES

1. Allbeck J. and Badler N. Toward Representing Agent Behaviors Modified by Personality and Emotion. In *Workshop Embodied conversational agents - let's specify and evaluate them!* AAMAS, (Bologna, 2002).
2. André E, Rist T and Müller J. Guiding the user through dynamically generated hypermedia presentations with a life-like character. In *Proceedings of the 3rd international conference on Intelligent user interfaces*, (San Francisco, California, United States, 1998), 21-28.
3. Cavazza M, Charles F and Mead SJ. Character-Based Interactive Storytelling. *IEEE Intelligent Systems* (2002), 17(4): 17-24.
4. Darwin CR. *The expression of emotions in man and animals*. Murray, London, 1872.
5. De Carolis B, Pelachaud C, Poggi I and Steedman M. APMML, a Markup Language for Believable Behavior Generation. In H Prendinger, M Ishizuka (eds). *Life-like Characters. Tools, Affective Functions and Applications*, Springer, 2003, 65-86.
6. Georg G and Jaulent M-C. A Document Engineering Environment for Clinical Guidelines. In *Proceedings of the 2007 ACM Symposium on Document Engineering*, (Winnipeg, Manitoba, Canada, 2007), ACM Press, New York NY, USA, 69-78.
7. Gratch J, Rickel J, André E, Cassell J, Petajan E and Badler N. Creating Interactive Virtual Humans: Some Assembly Required. *IEEE Intelligent Systems* (2002), 54-63.
8. Hoorn J and Konijn E. Personification: Crossover between Metaphor and Fictional Character in Computer Mediated Communication. In *The annual meeting of the International Communication Association*. (San Diego, CA, 2003).
9. Patel V, Arocha J, Diermeier M, How J and Mottur-Pilson C. Cognitive psychological studies of representation and use of clinical practice guidelines. *Int J Med Inf.* (2001), 63 (3);147-167.
10. Pelachaud C. Multimodal expressive embodied conversational agent. In *ACM Multimedia, Brave New Topics session*, (Singapore, 2005), 683-689.
11. Poggi I. Mind Markers. In M Rector, I Poggi, N Trigo eds. *Gestures, Meaning and use*, University Fernando Pessoa Press, Oporto, Portugal, 2003, 203-207.
12. Poggi I and Pelachaud C. Performative faces. *Speech Communication* (1998), 26, 5-21.
13. Rist T, André E, Baldes S, Gebhard P, Klesen M, Kipp M, Rist P and Schmitt M. A Review of the Development of Embodied Presentation Agents and Their Application Fields. In H Prendinger, M Ishizuka (eds). *Life-Like Characters: Tools, Affective Functions, and Applications*, Springer, 2003, 377-404.
14. Steedman M. *The syntactic process*. MIT Press, Cambridge, MA, 2000.