IADIS International Journal on WWW/Internet
Vol. 14, No. 1, pp. 1-14
ISSN: 1645-7641

THE ROLE OF A COGNITIVE BASED MODEL IN MULTIMODAL INTERACTION SYSTEMS DIALOGUE MANAGEMENT

Jonathan S. Prates, Sandro J. Rigo, Cristiano A. Costa and Jorge. L. V. Barbosa
University of Vale do Rio dos Sinos (UNISINOS), São Leopoldo, RS, Brazil

ABSTRACT

Researchers, in Multimodal Interaction Systems, devote substantial effort to the integration of external stimulus and signal, to the internal representation of this information and to the response generation. Nevertheless, they focus less effort on how to integrate studies in Cognitive Psychology regarding the human interaction subject. Therefore, it becomes interesting to evaluate the assumption that a dialog model based on known human factors should outcome better perception of the interaction process by the users. This paper presents an experimental approach to implement multimodal interaction systems where the main innovative aspect is related to the dialogue managing, that is based on the working memory model proposed by Baddeley and Hitch. A computational model was proposed and two different prototypes were built. The obtained results were found to be positive and it indicates relevant flexibility aspects together with anencouraging feedback from users.

KEYWORDS

Multimodal interaction systems, dialogue management, semantic web.

1. INTRODUCTION

The dialogue established between users and Multimodal Interaction Systems is a central element to provide a feeling of greater usability (Dumas, 2009; Bui, 2006). Dialogue management involves various activities, such as the representation of topics treated in the conversation, the choice of response alternatives, along with task and user models dealing. The demand for the creation of models providing more natural results (Jaimes, 2005; Cutugno, 2012) is related with a better treatment of natural language interaction aspects, but also with the challenges represented by the recent inclusion of body language facets in this context. This ability to expand the dialogue elements, considering not just words, is a challenge for Multimodal Interaction Systems, in particular due to the current availability of devices that make possible to

capture these very diverse data, such as data about user body (Hoste, 2011), aspects of facial expression and even brainwave-based elements (Leeb, 2013).

In most conversation systems between human and computers, the users need to inform words in predefined formats or phrases that meet prerequisites to guarantee that recognition can be possible and the system performs correctly the expected activities. Thus, we can say that these so called natural user interface systems often are considered notuser-natural. Other systems have as their main characteristic the question-answering model, where in fact there is nota real dialogue, considered as an exchange of information between the parties, but instead is observed a search process using the terms informed as input values (Pereira & Rigo, 2013). These two situations can end up in user frustration and generate poor interaction. A possible line of action for improvements in this regard can be the use of cognitive psychology studies about the working memory and information integration (Helene, 2003; Neto et al. 2009).

Multimodal interaction systems allow a friendly use of computing systems. They allow users to receive information and indicate their needs with ease, supported by new interaction resources. In this context, the central element is the dialogue, established between users and these systems. The dialogue management of these systems involves various activities associated with the representation of subjects treated, possible answers, tasks model and users model treatment. In implementations for these approaches, some demands can be observed to approximate the results of the interactions by these systems of interaction in natural language. One possible line of action to obtain improvements in this aspect can be associated to the use of cognitive psychology studies on working memory and information integration.

This article presents results obtained with a dialogue model for multimodal interaction systems based on cognitive model about the working memory, described byBaddeley and Hitch (2000). It aims to provide conditions for the generation of dialog elements perceived by the user as closer to real situations of natural language dialogue between people. The model deals with contextual information and information about previous user actions on the basis of known cognitive psychology representation of working memory. Based on the literature, it is possible to observe that this approach is not frequent in other known works in this area (Tan, Duan & Inamura, 2012; Hoste 2011). This research also has as objective to propose a flexible model for the treatment of dialogue in multimodal Interaction systems, in order to provide possibilities to handle the diverse and even new elements of data input and output. The use of open protocols and the decoupling of components allow the model to be applied in multimodal interaction systems already in place and also to be extended to new input elements. This research presents studies that supported this proposal and the justification for the described model's description. At the end, results using two prototypes for the model's validation are also shown.

This paper is organized as follows. In section 2 are presented conceptual models on the working memory and its relationship with dialogues. Section 3 describes related work. The proposed model is described in section 4, along with implementation details. Evaluation aspects are described in section 5. Finally, section 6 presents the conclusions of the work.

2. WORKING MEMORY AND DIALOGUE CONTROL

The multimodality in the communication shows that information with the same meaning, many times, can be expressed in different form. In the early 20th century, books, paintings and artistic presentations showed a unique form of communication, raising theoretical reviews because each

form contained its own methods, its own assumptions and arguments. To break this paradigm, emerges a concept where multiple semiotic methods together can provide information with more quality because it's part of human nature to use various channels, at the same time, to communicate. Multimodal communication is important to build the meaning clearly, unambiguously, through information that go beyond speech or writing, expressed also through images, emotions and feelings that can build the meaning of such information (KRESS; LEEUWEN, 2001).

Humans express themselves through a language using words, spokenor written, but also through signs and body expression. The man is able to create sentences with these words, forming speech. For Linguistics, the speech represents a coordinated sequence of phrases, which is not limited only to the speak act, but also involves cultural and social aspects (POPPEL, 1989). The simple act of wear a certain outfit, for example, is already a way of communication generated by human being. During the speech, the issuer must use words so that the communication is carried out effectively. The parties involved must be in tune with the general context treated in the conversation, in order to that information to make sense.

During a dialogue, the memory is responsible for maintaining a stream in conversation, while preserving the clarity for the parties involved, based on speeches and arguments (GODOY, 2010). One of the elements studied in this area is the working memory. For psychology, there are some definitions on the working memory concepts and models. The Baddeley and Hitch (1974), the more accepted and studied in cognitive psychology, says that short-term memory is responsible for keeping information temporarily during its processing. This memory has one of its main features the fact that is limited, keeping only minor information related to the context that is being experienced by the person.

The memorycan be considered as a complex process that supports the maintenance of the aspects associated with consciousness (IZQUIERDO, 2011). She is an important cognitive component involved in understanding during communicating. These characteristics as fundamental to the dialogue between human beings are studied in the field of human-computer interaction and used as the basis for several works. This research analyzes the characteristics of the model proposed by Baddeley and Hitch (1974) and their use as part of the component responsible for dialog control in a Multimodal Interaction system. The approach was adopted in this work by suggesting the possibility of obtaining improvements in human-computer interaction, linked to the use of cognitive psychology studies on working memory and information integration.

The memory of humans has a very complex task and is one of the main elements responsible for the conscience of each individual. The memory is fundamental to locate information originated in the past and allow future decisions to be taken. It also contributes to the generation of the sense of continuity and clarity, so necessary for humans (Godoy, 2010). The working memory is responsible for reasoning, comprehension and learning. This memory component works maintaining a small information history. During a dialogue, this component is essential to keep in focus the information from previous speeches or interactions(Izquierdo, 2011).

The working memory model proposed by Baddeleyand Hitch (2000) claims that at least three support systems are responsible for short-term memory processing. Accordingly to Helene (2003) the model has a central executive component, responsible for coordinating support subsystems. The other components are three subsystems: *a)*VisuospatialSketchpad, capable of storing information acquired through images, such as colors, sizes and location of the given object; *b)*Phonological Loop, which features phonological cycle in order to avoid the loss of information, such as, for example, a phone number; *c)*Episodic buffer, a subsystem of limited

capacity that is able to pick up information from long-term memory. It is the task of the executive central to direct attention and to discard information no longer relevant. These three components represent the components dedicated to temporary retention of information, while they are associated with other systems of long-term information retention involving the language, episodic memory and visual memory. The figure 1 describes these elements and indicates the interations considered (HELENE; XAVIER, 2003).

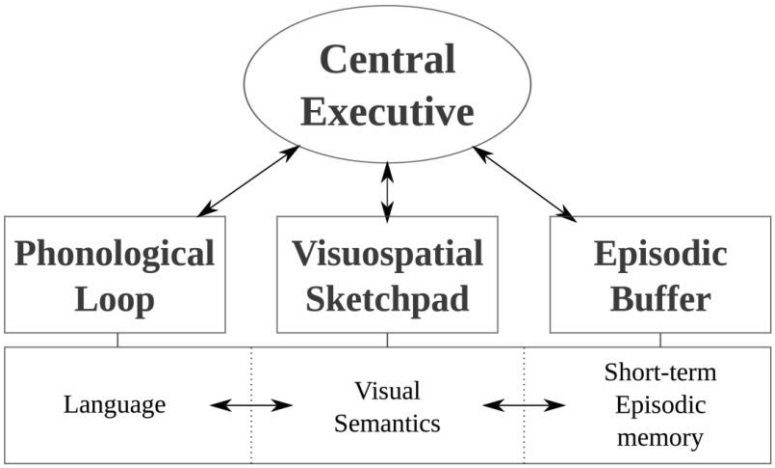

Figure 1. Working memory model by Baddeley and Hitch

The representations related to episodic memory are integrated in a manner relevant to language aspects. All the formation of the human language has more divisions than specific words and phrases. A speech is not composed of isolated information with loose sentences, but rather by a group of judgments directly related(Jurasky& Martin, 2000). Speech is a generic term used to define a group of sentences in a language, and can be one-way, known as monologue, or multidirectional, known as dialogue. Studies on elements and structure of discourse supports several works on conversational systems that present dialog managers (Rotaru, 2008). The dialog manager's role in the Groz and Sidner (1986) is to represent the linguistic context, the information interpreted the plan for the user and system responses. Previous interactions are stored in a template called dialog track. This component serves as a history of previous actions. Through the dialogue track it is possible to write information to restore the dialogue, because it stores all previous interactions.

3. RELATED WORK

This section describes some aspects of related works that were important in the definition of the work presented here, such as the search for flexibility in the interaction with new components and focus on management aspects of the dialogue. Some works incorporate these notions, as in the system COMIC: *COnversational Multimodal Interaction with Computers* (Pfleger, 2004), where the working memory is a representation of controlled objects that describe the situation in context. In the system described by Tan, Duan and Inamura (2012) the AIML (*Artificial*

Intelligence Markup Language)is used for dialogue generation. The system MIND: *Multimodal Interpreter for Natural Dialog* (Chai, 2005) has as a characteristic to use varied contexts (dialogue and domain, for example) to improve multimodal fusion of entries data. These features allow for a better interaction with the user. However the components of the model suggested in MIND are quite static, making difficult the expansion by third parties. Hoste (2011) presents a new approach to fusing multimodal inputs, allowing the use of such information at various levels, as well as a modeling approach that allows system modularization. This characteristic is fundamental to an expandable model, as proposed in the present research.

The work of Schroder (2010) is a modular work that allows interaction through messages using the XML standard. One of the advantages of this model is the use of standards recommended by W3C, allowing thus facilities for the system extension. As well as the model proposed in this research, the SEMAINE API uses a message-oriented middleware for sending messages between its components enabling decoupling and easy reuse of components. In the MATCH system (Johnston *et al*, 2002) the authors present an approach that allows modular development and fast multi-modal applications supported by context information. In spite of supporting user profiles, the system MATCH does not maintain a dialog history for future interactions.

We compared the related works described, regarding the way they provide support for: dialogue control management; devices and resources integration; data sources to be integrated in the dialogues; psychological based models use. As can noted in the related works, the advances in multi-sensorial perception devices use in the context of multimodal interfaces have been growing. At the same time, there is a gap to be filled regarding information utilization about the diverse user input. The proposed model was developed to fulfill this gap and to simulate aspects of working memory. None of the related works has the ability to generate a dialogue control with a human memory based model, such as the model proposed in this work. Similarly, none of the related works can easily integrate new components and devices in the multimodal fusion process, nor can they use external data sources to enrich the dialogue interaction.

4. PROPOSED MODEL

The dialogue management component based on a cognitive model is the differential presented by this work. This component reuses data from previous interactions in order to drive and support the next dialogue interaction. We termed this component as a working memory model. The elements of our working memory model are the main point of dialogue generation. Also, we present a flexible method to handle distinct kinds of input data and output representation. In order to make easier and promote the interaction with other input devices, the proposed model uses the well-known format EMMA[1]. This format has been developed by W3C[2]in order to define a pattern for this kind of data exchange. Different types of data can be used in this model, such as strings or text information, speech, gestures and signs, context and environment and user's location as well.

The proposed model was divided into three macro steps. Figure 2 shows a model's architecture overview and its components: input, control and output. This figure uses the

[1]http://www.w3.org/TR/emma
[2]http://www.w3.org/

Technical Architecture Modeling[3] (TAM) format. As described before, the input component was designed to accept data from different mechanisms and devices. Each external input device can be treated in accordance with a specific processing system for that data, by using EMMA format. In this work, we used a AMQP-based message-oriented middleware[4] (MOM). Control component has the elements for data fusion and dialogue control. Also, this component has some auxiliary functions to support these elements. The output component has a module responsible to interpret and integrate the output information. This component drives this data to correct output device. All data exchange is performed by the message oriented middleware. The fusion component uses data from different devices, in order to infer a semantic interpretation. The evaluation of each received data by system will be mixed with other input data to identification of purpose and context.

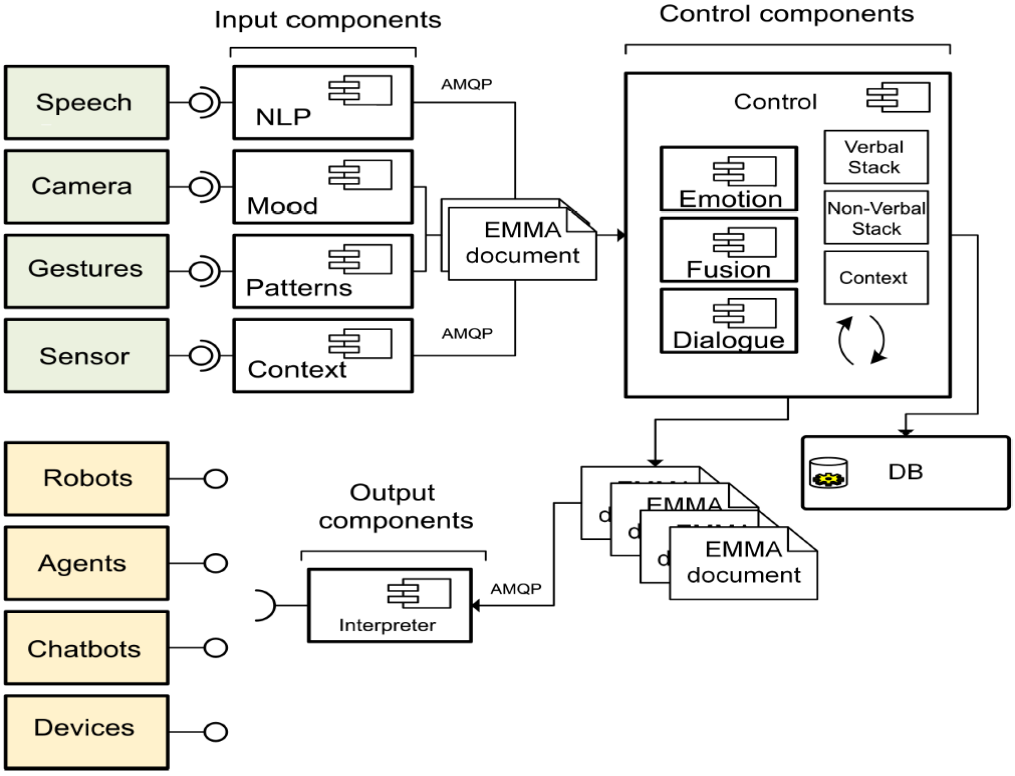

Figure 2. Model's architecture overview

Similarly to the general architecture Dumas, Lalanne and Oviatt (2009), which shows that multimodal interaction systems have a component responsible for informationintegration, in the proposed model the component responsible for integrating it (the control component) is divided into a fusion module of multimodal inputs, a connected data access component, the dialog manager, verbal memory, nonverbal memory and the environment and context information. As a

[3]http://www.fmc-modeling.org/fmc-and-tam
[4]http://www.rabbitmq.com

differential in relation to the work of Dumas, Lalanne and Oviatt (2009), stands out: the dialog manager, responsible for the interpretation of the inputs and the control of the topic in attention; the elements of verbal and nonverbal memory that stores a history on previous interactions; the fusion components and connected data access. These aspects are represented in Figure 2 control component and were defined in order to represent the cognitive model proposed by Baddeley and Hitch (1974).

Due to many differences between types of user inputs, the fusion module receives this information from some devices and performs its unification. For example, the system can receive the geographic coordinate information and will try to discover the user's location. This data will be used together with user speech input. Thus, queries to linked open database, used to bring information regarding the dialogue topic, can be more specific through incorporation of geographic information. The fusion module considers three main types of data, which are addressed through the use of EMMA format. These types are the following:*a)* Environment and context: Data classified as environment and context are often forgotten during a dialogue by humans, but they are perceived and understood unconsciously, for example, the temperature in degrees and the location (latitude and longitude); *b)* Dialogue: Data classified as dialogue refer to written or spoken inputs. This kind of input is the base of conversation between computer and user;*c)* Expression or emotion: Data classified as expression or emotion information refer to facial or body gestures that can change the direction or add information to the dialogue.

The dialogue management module is the main point of the proposed work. It is responsible for interaction process with users. This module has the responsibility to keep the conversation in context, by using previous information from fusion module. For dialog generation, the model uses its working memory, which keeps the history of interaction and will help the system to identify the topic of conversation. The dialogue management module is responsible for the subject interpretation after multimodal fusion takes place and is also dependent on the long term memory representation. At the moment the user change the subject, is performed a query to a knowledge base. This base serves for the model as a long-term memory, where the system will rescue data about a particular subject. In this work, we used the DBPedia, which has a large collection of information. Other more specific related databases could be used, reducing the scope and increasing accuracy.

The working memory stores useful data and use them to create richer and insightful queries. To improve the recognition of the meaning user's input, the model will use its working memory and some auxiliary functions from natural language processing. For example, given an input, the system can use previously stored data to create queries and search for information in knowledge bases. To find the meaning of certain information, the model records in its working memory this new interaction and returns an answer action for the user. Knowledge bases are defined in the model as a layer of access to linked data, serving as a foundation for the identification of user input. This approach allows a gradual expansion for the system. The output component main function is the integration and routing information generated by the dialog module. Thus, this component can be adapted to the needs of each multimodal interaction system, allowing more kinds of data output.

4.1 Dialogue Management

The objective of the dialogue management module is to maintain the registry and to use previous information to provide more natural user interaction. Figure 2 summarizes the function of this

module, from elements representing the input and output of data. Each component sends a message according to their function. The contents are sending in EMMA standard. Given a particular input through one of the system components, it will be interpreted and sent to the system. For example, a component responsible for speech recognition will send this information as text. Whereas, a component responsible for emotion recognition can get pictures from a camera, infer the user's emotional state and send this information to the system. These messages are received asynchronously and its values are recorded in the working memory.

The topic of conversation is important for managing the dialogue and set the point of attention in each step of the conversation. When the conversation attention point is changed, the system will query an open linked database to recover data about this topic. This information is stored in working memory and generates responses to possible user questions. In the case of implicit interactions (gestures or expressions), the model uses the non-verbal memory to store this information. This data will be taken to identify user location, current emotional state and other attributes that can have influence on the dialogue.

In addition to the elements described by figure 3, the management of the dialogue makes use of an internal component for the treatment of text messages, which uses natural language processing capabilities to identify the main information from each phrase, required to find correct question type.This implementation was based on a study of Li and Roth and adapted to Portuguese (Li and Roth, 2006). Figure 3 shows that thisprocess occurs in block "identify main topic".

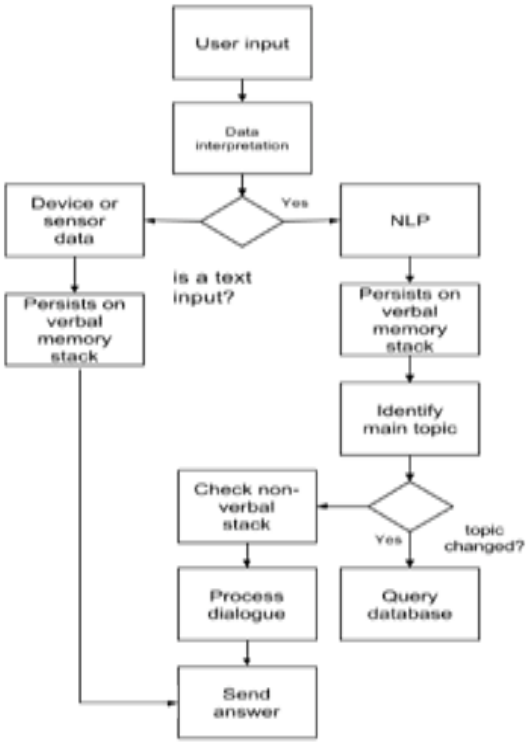

Figure 3. Dialogue flowchart

Finally, the management of dialogue internally uses a finite state automaton (with predetermined states), which allows system to control the dialogue states in an efficient way. Operation of this finite automaton is presented in figure 4. The purpose of this component is to prevent processing user input incorrectly. Also, the dialogue management allows questions like "What is the weather forecast for today? ". It can be answered using data from non-verbal memory, based on location obtained by GPS device, for example. Supported states are as follow:*START* - initial state of the machine; *OPENING*- input state for dialogue; *REQUEST/RESPONSE* - user did a direct question;*REQUEST/CLARIFICATION* - user asks a question related to context or previous question;*CONFIRM* - user accepts the answer;*REJECT* - user does not accept the answer;*CLOSING* - the user finishesthe dialogue.

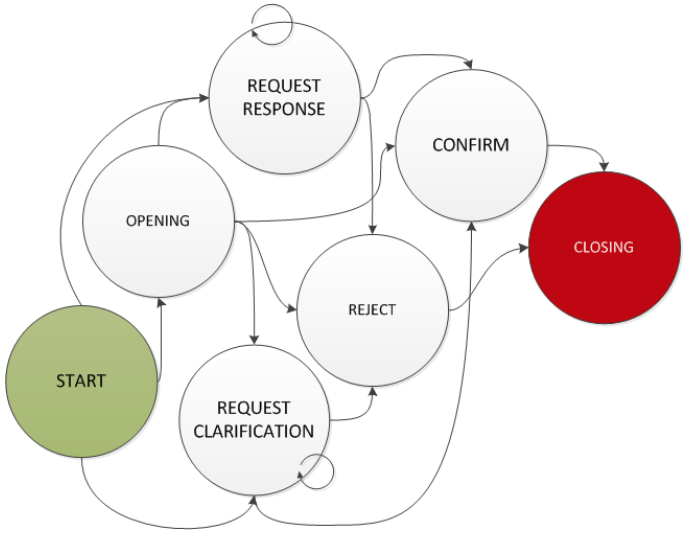

Figure 4. Dialogue management support

4.2 Implementation Aspects

To demonstrate and assess the feasibility of integrating the items mentioned in the proposed model, two prototypes were created. The former is a program of questions and answers, based on first one interaction through messages in text format. Its main focus is on the validation of the elements involved in the exchange of messages, allowing the loose coupling between components. This prototype also allows the use validation of EMMA format in the data representation, AIML processor and auxiliary libraries and services of natural language processing.

The second prototype is an application for smart phones, where you can observe aspects of multimodal interaction through emotions, sensors and speech recognition. Both prototypes are intended to highlight the items involved in managing the dialogue, the use of environmental information and the use of database linked to answer questions from users. To enable the integration of data in different formats originating from different devices, we used the standard EMMA, described above, as it ensures a perspective of compatibility and interoperability. The

integration between prototype and its modules is performed through a MOM, allowing complete independence between different input devices and system components.

For dialogue generation we used an AIML processor based on the Python language, the PyAIML library[5]. During the interaction with the user, the system consults a linked open database. DBPedia was used in this work (Auer *etal*, 2007). The query results, returned in JSON-LD[6]format, are used in AIML generation contexts. The processing of the textual input is performed by NLTK[7]by generating morph-syntacticinformation and AlchemyAPI[8]service to perform the natural language processing, subject's identification and sentiment analysis.

The second prototype implementation has some additional features that are not present in the first prototype. This prototype is a question-answer application for smart phones, with data integration arising from the camera and the GPS sensor. Furthermore, the prototype performs speech recognition to treat input and voice output to output data held by the GUI. This application has been divided into two parts for its operation due to processing limitations and development on the mobile platform. The code developed in this work is availableat https://github.com/.

5. OBTAINED RESULTS AND DISCUSSION

The evaluation of the model proposed in this work happened in two stages. The first is the adoption of the evaluation by scenarios. The goal here is to list and highlight the main features of the model in operation from experiments conducted on the prototypes described. The main focus of this step is to evaluate the dialogue management and your working memory in different situations. The second evaluation stage has as main objective to validate the proposed dialogue management comparing interactions generated by the system with real users expected result.

For the first stage of evaluation, the following scenarios were used to highlight the main features of the proposed model. The first scenario tested involves the response of several questions on the same topic, whose information is accessed from data obtained in DBPedia. In the second scenario the system maintains a dialogue, but identifies the change of topic and manage to continue the dialogue with the user with the changes subject being the conversation focus. A third scenario demonstrated the use of the proposed model combining textual entries and emotions.

An example of obtained results in the second scenario is illustrated in Figure 5 below, where can be observed the subject change from the initial subject and the correct response of the system. In the second evaluation step it was proposed to a group of 37voluntaries that write some direct questions to the system to respond, as well as the answer they expected for those questions. The questions raised by these users were used in interaction sessions with the second prototype and then returned with the answers generated by the system. These responses were compared to expected answers for each user. This experiment was important to know how people make their issues, in general, thus creating a more effective testing scenario, avoiding delays due to availability of each participant and focus on validation of the model instead of the prototype.

[5]http://pyaiml.sourceforge.net/
[6]http://json-ld.org/
[7]http://www.nltk.org/
[8]http://www.alchemyapi.com/

```
User: Hello!
System: Hello!
User: What do you like to eat?
System: Churrasco
User: And what is that?
System: Churrasco is a Portuguese and Spanish
term referring to beef or grilled meat more
generally, differing across Latin America and
Europe, but a prominent feature in the cuisine
of Brazil.
User: What ingredients are used to make a
Churrasco?
System: Typical Brazilian Churrasco uses
chicken heart, sausages, bread with garlic
sauce, sliced picanha and chicken legs.
User: Ok. Who was the director of Titanic movie?
System: James Cameron
User: What is the movie synopsis?
System: Titanic is a 1997 American film written
and directed by James Cameron. It is a
ficccionalizada story of the sinking of RMS
Titanic (...)
```

Figure 5. Example of results in test scenario

People participating in this experiment were selected by convenience and have a suitable profile for the experiment. The age range varies between 22 and 57 years. The professions of the participants are predominantly in the areas of information technology, administration, health and communication. The guests for the experiment received instructions to fill out a form with the elements needed to perform the test.

Each user was asked to indicate a data set that allows you to validate if the prototype performs its operation according to the expected, namely, detecting the current context and reply to a related question. In this way, participants indicated an initial question and its answer, in addition to a second question about the general context of the first one and of the expected response. An example of this data set that illustrates this activity can be exemplified with the question "What's the capital of Norway?", being that the answer expected by the participant was "Oslo". The second question was "How many people it has?", to which the expected answer "593 thousand inhabitants". Should be noted that the second question do not mention the entity that was stated in the first question and it is expected the system to deal with this situation.

We have defined four main criteria to get our results. Through a Likert scale, we were able to measure the quality of features described in this work. The first criterion was defined as "topic identification". It means that the user should choose a subject area from a given range (like political, cinema, food etc), and ask a question about it. For this case, the system must be able to identify this subject, by parsing its input, and reply which subject this user are talking about. This predefined subject range was motivated by limitations of natural language processing in Portuguese. It was needed to find a more robust and smarter approach to handle Portuguese sentences, such as *"dialog act tagging*" and other. Implementing searching by relevance using techniques like TF-IDF or more comprehensive algorithms might prevent errors while trying to find topics in DBPedia, as suggested in *"Enabling Keyword Search on Linked Data*

Repositories: An Ontology-Based Approach" (Bobed *etal*, 2013). It is important to remember that the main goal of this interaction is to evaluate the model proposed on this work, not the prototype itself.

In second criterion, defined as “Keeps the context”, it was asked to each user to perform a question linked to result of first question, without any direct mention of the main subject. We could verify whether the system was able or not to keep the context correctly. Therefore, if the subject changes, the system, should be able to answer to handle this new context. The third criterion was defined as “Correct answering”. It aims to compare the sentences generated by the system with the sentences expected by users. The last criterion was defined as “Sentiment analyses”. The users were asked to send two simple phases. The first one should be “negative” whilst the second one should be positive.

It is possible to see that, in the most of cases, the main subject’s topic from each question was identified properly. We have noticed import points since the success of the system depends on other components, such as DBPedia and natural language parser. For instance, one user asked “Quantosmeses tem um ano?” - in english - “How many months has a year?” and “Qual o mês do Natal?” - in english - “What is christmas’ month?”. System couldn’t tag “Ano” - year in english- as a named-entity (expression of time). Also, system wasn’t able to link 1 year between 12 months and christmas. There is no link between this meanings in DBPedia. Thus, the result of this question was considered negative by user, however, we understood that it should be improved in natural language processing component and DBPedia in Portuguese.

Another user performed the follow questions: “Quemescreveu O Capital?” - in english - “Who wrote O Capital?” and “Qualsua data de falecimento?” - in english - “What is he death date?”. The system wasn’t able to recognise “O Capital” as a named-entity (in Portuguese) or DBPedia entity. Nevertheless, by citing Karl Marx in a test case, the system was able to answer the correctly. The same problem happened with questions: “Ondeestão Scooby-Doo e Salsicha?” - in english - “Where are Scooby-Doo and Shaggy?” and “O queelesprocuram?” - in english - “What are they looking for?”. The system skipped Scooby-Doo as a named-entity and the second was a subjective question and isn’t in scope of this work.

It is possible to see that each user has a different manner to ask a question, even following basic instructions. We noticed also when the topic is recognised correctly, the model proposed in this work might answer user’s question keeping the main context of a conversation.

Finally, the evaluations carried out from questions proposed by users identified a positive perception with respect to the context of the messages, in more that 76% of the cases. These evaluations also made it possible to see the dependency of prototypes as to its components, in particular the components used to the knowledge base and the processes of natural language processing. In cases where these components do not have the desired information or who fail to properly process messages, the result is not perceived in a satisfactory manner. It is possible to observe these aspects in other systems in that the confidence on specific knowledge of an area can be decisive for the outcome of the multimodal interaction system.

6. CONCLUSION AND FUTURE WORK

In this work were exploited improvements in the control of dialogues in Multimodal Interaction systems with the use of known characteristics of working memory models. The objective was to identify aspects that can act in the best generation of dialog formats, with the maintenance and

processing of information that describe the memory of dialogue and to include in the general context of the new dialogue events related with the previous information. In addition, the model simulates the use of long-term memory, from searches performed in linked databases or knowledge bases, complementing the information available from the dialogue and allowing the system to present aspects of greater flexibility.

The use of this type of tool can support more autonomous and creative activities of users in diverse situations, with facilitated interaction generated and data access made available in online format. The proposed model, considering the related works, presents as a differential the use of these aspects of working memory and dialogue context in an integrated manner, to support the generation of more flexible dialogues with most suitable perception on the part of the user. The validations performed to identify the possibility of generating appropriate contexts for dialogues with use of multimodal information were positive in their major extent.

As further work we intent to implement more broadly some of the tests conducted, involving greater amount of input and output devices of data, in order to assess how the model developed behaves in the treatment of most aspects of the dialogue. In this work, the working memory only serves to interact within a dialogue, not being used to store user profile data and but future activities will extend the model for such operation.

REFERENCES

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., & Ives, Z. (2007).DBPedia: A nucleus for a web of open data (pp. 722-735). Springer Berlin Heidelberg.

Baddeley, A. D., & Hitch, G. (1974). Working memory. Psychology of learning and motivation, 8, 47-89.

Bobed, C. Esteban, G., Mena E., "Enabling Keyword Search on Linked Data Repositories: An Ontology-Based Approach" (2013). International Journal of Knowledge-based and Intelligent Engineering Systems, ISSN1327-2314, volume 17, number 1, pp. 67-77, 5.

Chai, J. Y., Pan, S., &Zhou, M. X. (2005). Mind: A context-based multimodal interpretation framework in conversational systems. In Advances in Natural Multimodal Dialogue Systems (pp. 265-285). Springer Netherlands.

Dumas, B., Lalanne, D., &Oviatt, S. (2009). Multimodal interfaces: A survey of principles, models and frameworks. In Human Machine Interaction (pp. 3-26).Springer Berlin Heidelberg.

GODOY, J. P. M. C. Integração de informaçõesvisuais e verbaisnamemória de trabalho.

2010. Dissertação (MestradoemCiência da Computação) — Universidade de São Paulo,

2010.

Grosz, Barbara J., and Candace L. Sidner. "Attention, intentions, and the structure of discourse." Computational linguistics 12.3 (1986): 175-204.

Hoste, L., Dumas, B., &Signer, B. (2011). Mudra: a unified multimodal interaction framework. In Proceedings of the 13th international conference on multimodal interfaces (pp. 97-104). ACM.

Helene, A. F.; Xavier, G. F.A construcao da atencao a partir da memoria.Revista

Brasileira de Psiquiatria, [S.l.], v. 25, p. 12 – 20, 12 2003.

Jaimes, A., &Sebe, N. (2007). Multimodal human–computer interaction: A survey. Computer vision and image understanding, 108(1), 116-134.

Johnston, M., et al. (2002) "MATCH: An architecture for multimodal dialogue systems." Proceedings of the 40th Annual Meeting on Association for Computational Linguistics.Association for Computational Linguistics.

Johnston, M., et al. (2009). Emma: Extensible multimodal annotation markup language. World Wide Web Consortium Recommendation REC-emma-2009021.Technical report, W3C.

KRESS, G.; LEEUWEN, T. van. Multimodal Discourse: the modes and media of

contemporary communication. 1. ed. New York: Oxford University Press, 2001.

Neto, A. T., Bittar, T. J., Fortes, R. P., & Felizardo, K. (2009). Developing and evaluating web multimodal interfaces-a case study with usability principles.In Proceedings of the 2009 ACM symposium on Applied Computing (pp. 116-120). ACM.

Oviatt, S., Coulston, R., & Lunsford, R. (2004). When do we interact multimodality?: cognitive load and multimodal communication patterns. In Proceedings of the 6th international conference on Multimodal interfaces (pp. 129-136). ACM.

Picazzo, F. L., Guzmán, F. V., Aguilar, S. F., & Parada, B. S. (2014). EMINUS SYSTEM OF EDUCATION DISTRIBUTED IN SUPPORT OF MULTIMODAL EDUCATION. INTED2014 Proceedings, 4488-4494.

POPPEL, E. Fronteiras da Consciência - A Realidade e aExperiência do Mundo. [S.l.]:

Edições 70, 1989.

Schröder, M. (2010). The SEMAINE API: towards a standards-based framework for building emotion-oriented systems. Advances in human-computer interaction, 2010, 2.

Turk, M. (2014). Multimodal interaction: A review. Pattern Recognition Letters, 36, 189-195.

Vinoski, S. (2006). Advanced message queuing protocol. IEEE Internet Computing, 10(6), 87-89.