Vol. 4, No. 1 Jan 2013 ISSN 2079-8407
Journal of Emerging Trends in Computing and Information Sciences

http://www.cisjournal.org

Multimodal Fission For Interaction Architecture

[1] Atef Zaguia, [2] Ahmad Wahbi, [3]Chakib Tadj, [4]Amar Ramdane-Cherif
[1,2,3] MMS Laboratory, Université du Québec, École de technologie supérieure
1100, rue Notre-Dame Ouest, Montréal, Québec, H3C 1K3 Canada
[2,4] LISV Laboratory, Université de Versailles-Saint-Quentin-en-Yvelines France

[1] atef.zaguia.1@ens.etsmtl.ca, [2] ahmad.wehbi.1@ens.etsmtl.ca, [3] ctadj@ele.etsmtl.ca, [4] rca@prism.uvsq.fr

ABSTRACT

Since the eighties, the rapid development in the world of information technology has made possible to create systems that interfere with the user in a harmonious manner. This is due to the emergence of a technology known as *multimodal interaction*. This technology allows the user to use natural modalities (speech, gesture, eye gaze, etc.) to interact with the machine in a richer computing environment. These are called *multimodal systems*. These systems represent a remarkable deviation from using conventional systems, such as windows-icons, to a human-machine interaction, providing to the user more naturalness, flexibility and portability. Generally, these systems integrate multimodal interface in input and output. Via the output interface, the system should be able to choose among the available modalities, those that meet the best environmental constraints. It should also be able to interpret a complex command and divide it into elementary sub-tasks and present them in the output modalities: it is called *multimodal fission*. Our work specifies and develops fission components for a multimodal interaction and presents an effective fission algorithm using patterns, when various output modalities (audio, display, Braille, etc.) are available to the user.

Keywords: *Multimodal fission, pattern, human-computer interaction.*

1. INTRODUCTION

Since the advent of computers, one of the biggest challenges in informatics has always been the creation of intelligent systems that enable transparency and flexibility of human-machine/machine-machine interaction(Sears et Jacko, 2007)(Alm, Alfredson et Ohlsson, 2009).

In our days, computers and intelligent systems have become increasingly ubiquitous. Users are increasingly hungry to systems that are easy to use and intelligent.

Communication plays a primordial role in our common life. It allows humans to understand each other and to be connected as individuals or as independent groups. This communication is done across several natural modalities as speech, gestures, gaze and facial expressions.

Humans have a highly developed ability to transmit ideas between each other and to react in an appropriate way. It is owed by the way of the shared of the language, as well in the common understanding of the functioning of things and an implicit understanding of daily situations.

But without external intervention, machines do not understand our language, do not understand how the world works and cannot collect information about a given situation. Researchers aim to always satisfy user's needs and provide systems that are smarter, more natural and easier to use.

Therefore various efforts have been geared toward the creation of systems that facilitate the communication between humans and machines and to allow to the user to make use of multimedia devices relying on natural modalities (sight, speech, gesture, etc) to communicate or exchange information with applications, machines, etc.

These systems receive inputs from the sensors and gadgets (camera, microphone, etc.) and they make the interpretation and understanding of these inputs or multimodalities. A known example of such systems is Bolt system (Bolt, 1980) "Put that there" where it used the gesture and speech to move objects. These systems generally comprise a multimodal input interface and amultimodaloutput interface. Via the output interface, the system should be able to choose among the available modalities, those that satisfy the best of environmental constraints, functional requirements of the task and user preferences. Several multimodal applications have been created ((Beinhauer et Hipp, 2009), (Caschera et al., 2009)(Robert, Khaled et Tharam, 2005)).These systems represent an effective solution for users, particularly for those who cannot use a keyboard or a mouse, the partially sighted users, the users who are equipped with mobile apparatuses, the users who are disabled, etc, to use their natural modalities (word, gesture, look, etc) to interact with the machine with a richer and more various expressiveness what they call the multimodal systems. Therefore the multimodal systems ameliorate accessibility for different users.

Each multimodal system (Meng et al., 2009) is based on tow essential components (Fig 1): fusion ((Zaguia et al., 2010b), (Wehbi et al., 2011), (Atrey et al., 2010)) and fission ((Costa et Duarte, 2011), (Ertl, Falb et Kaindl, 2010)).

Vol. 4, No. 1 Jan 2013 ISSN 2079-8407
Journal of Emerging Trends in Computing and Information Sciences

http://www.cisjournal.org

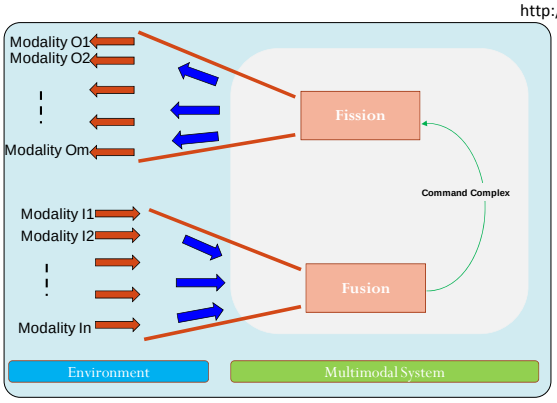


Fig 1: General architecture of multimodal system.

Fusion usually refers to a process combining events at the entrance to understand request of a user in his environment and to achieve a single but complex command.

Using the fission process, the system interprets a complex command, divides it into elementary sub-tasks and presents these sub-tasks as elementary output modalities. For instance, if a robot receives a command "Move an object A to a position (X,Y)", the system subdivides this complex command to:

a. Move toward "A" using mobility mechanism,
b. Take "A"using manipulator,
c. Head to position (XY) using mobility mechanism,
d. Drop "A" using manipulator.

The mobility mechanism and manipulator are services related to the output modalities.

In our work, we focus specially on 1) the services connected to the output: Multimodal Fission and 2) the creation of multimodal interaction system.

The rest of this paper is organized as follows. Section II presents ours research problematic. Section III takes note of other research works that are related to ours. Section IV discuses the modalities selection and multimodal fission system. Section V presents a simulation for a scenario. The paper is concluded in section VI.

2. CHALLENGES AND PROPOSED SOLUTION

Our main objective is to develop an expert system, capable of providing services to different multimodal applications.

The system receives a complex command, subdivides it into elementary sub-tasks and presents them to output modalities. Here, we enumerate some challenges that need to be addressed and the proposed solutions in order to develop our system.

a. What are the modules required for the architecture of a fission system? Here we will specify, define and develop all necessary components of the system. We will also show how they communicate.
b. How do we represent the multimodal information? We will model semantically the environment. To achieve our goals around the fission component, we create an context-sensitive architecture able to manage multiple distributed modules and automatically adapts to dynamic changes of the context of interaction (user, environment, system) (Zaguia et al., 2010a).
c. How do we perform the fission process? We will introduce an algorithm that describes the fission mechanism. It includes the rules of fission and the rules of selection of output modalities.
d. The fourth problem how to validate our architecture (formalism)? We focus more closely on the design, specification, construction and evaluation of our fission architecture.

3. RELATED WORK

In general, modality refers to the mode of interaction between man and machine for input and output data. With the use of traditional computing, human-computer interaction is limited a traditional mouse, keyboard and screen.

Multimodality is an effective solution that enriches the human-machine communication. It allows a) a more flexible interaction between the user and the machine and b) a use of natural modalities to interact with the machines.

Current researches try to find solutions to subdivide the complex commands into elementary sub-tasks and present them to the output modalities. Many studies have addressed this problem. The system presented in (Benoit et al., 2009) is a multimodal system "driver simulator". This system has data video and biological signal as inputs and audio sound, visual messages and wheel vibration as outputs. The system reacts, in real time, to the situations of fatigue and stress of the driver. In (Foster, 2005), the author presents a system commonly used in the construction of houses and especially the design of bathrooms. The interface of this multimodal system includes in input speech and pen. They are used in an intuitive and integrated way. The feedback (output) is generated by voice, graphics and facial expressions of the talking head (Foster, 2005). After the design, there will be a 3D visualization. The interaction between the user and the system is designed in the way that the system supports the user in a continuous manner during the design of the bathroom. In (Poller et Tschernomas, 2006),SmartKom is a multimodal system that combines the inputs speech, gesture and biometrics and manage the outputs speech, gesture and graphics. This system provides a visual display that includes the

natural language text and a speaking avatar. Throughout the use of the system, the user gets a consistent and enjoyable experience through personalized interaction agent (avatar), called Smartakus. In the paper the authors present three different scenarios of SmartKom applications:

- SmartKomPUBLIC: This scenario represents a multimodal kiosk. The user can use the system to scan objects, send emails, make calls, etc.
- SmartKomHOUSE: The system acts as an intelligent information system at home. The user is equipped with a tablet which controls the television, can have access information about a given TV station programs, record programs etc. while using natural methods.
- SmartKomMOBILE: The system behaves as a tour guide or a GPS navigator for a user with a PDA.

Most multimodal systems studied in the literature use only two modalities and their application-specific architectures are targeted. So most of the multimodal systems studied focus on the merger, but no advantage for fission. In some cases, they use a static database or in other cases, they use predefined scenarios to achieve fission.

In (Benoit et al., 2009), the proposed fission system is very simple. They test if the values entered are in some intervals and through this test, the system will generate alerts. In our case, we have a complex command and the goal of the fission module is to subdivide it into sub-tasks. Every elementary task will be presented in the available output and adequate modality (ies).

4. MODALITIES SELECTION AND MULTIMODAL FISSION SYSTEM

Before presenting our work, we present the definitions found on multimodal fission in the literature. Poller(Poller et Tschernomas, 2006), defined the fission as "the partitioning of a presentation into tasks for different media, called multimedia fission". Foster (Foster, 2002) defined it as the process of realising an abstract message through the output based on the combination of available modalities. For Landragin(Landragin, 2007), multimodal fission is related to the distribution of information across multiple modalities. The main role of the multimodal fission is to determine which message will be generated with each modality. The objective of multimodal fission is to move from an independent presentation of modalities to a coordinated and coherent multimodal presentation.
In general, a fission process is presented in Fig 2. It consists of 3 essentials modules:

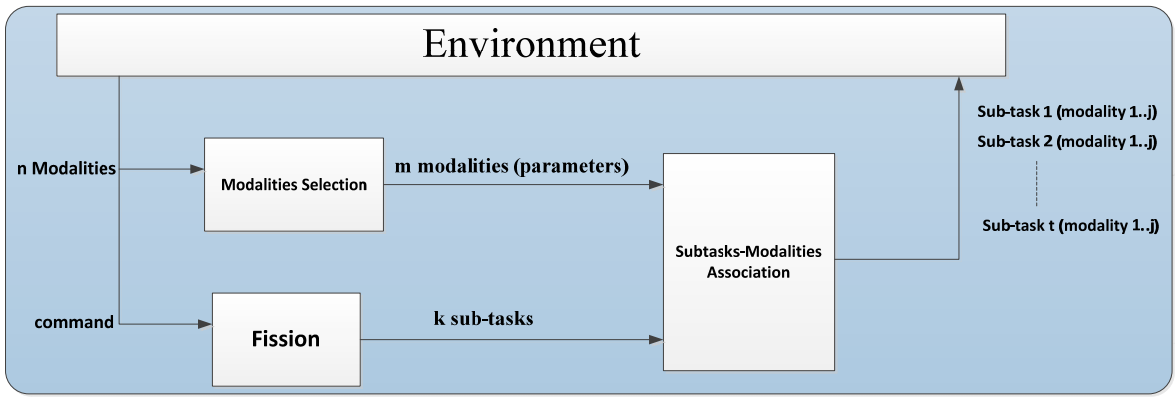


Fig 2: General view of the fission process.

- « Modalities selection »: The purpose of this module is to select which modalities can be used according to the state of the environment. This module is detailed in section B.
- « Fission »: performs the fission as described earlier. The input of this module is the command and the output the elementary subtasks.
- « Subtasks-Modalities Association»: its role to associate for each subtask (output for fission module) to the modalities selected by the module « Modalities selection ». This module is detailed in section C.1.1.

a. Multimodal Fission Architecture

In this section, we describe the proposed approach and the modules involved in the design and implementation of the architecture of the multimodal fission. The architecture is able to interact with multiple applications as shown in Fig 3.

Vol. 4, No. 1 Jan 2013 ISSN 2079-8407
Journal of Emerging Trends in Computing and Information Sciences

http://www.cisjournal.org

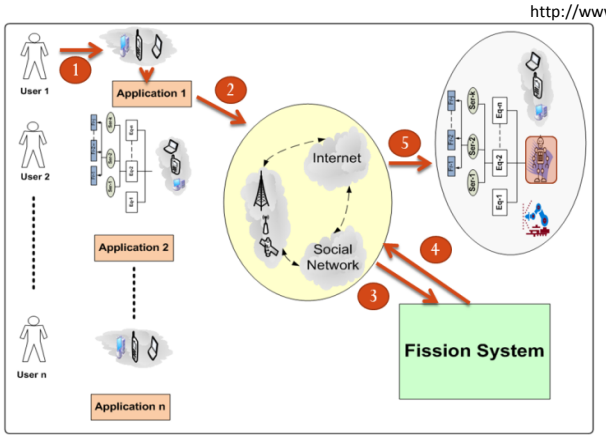


Fig 3: Interaction of our system with several applications.

For instance, suppose Application 1 is a system that controls a remote robot. As shown in Fig 3, the user generates a command using a phone or a computer (stage 1), the application (stage 2) sends the command via internet or social network to our multimodal system. The system receives the command (Stage 3), performs the fission and sends the elementary sub-tasks to the robot (Stage 4 and 5).

The proposed system architecture illustrated in Fig 4is modular and distributed. It contains five main modules (Fig 4).

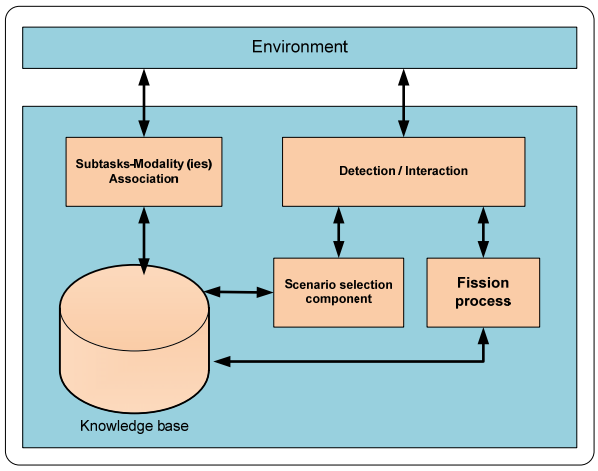


Fig 4: Multimodal fission system architecture.

- Detection/Interaction: This module will interact with the environment to allow modules "Fission process" and "Scenario selection component" to achieve fission. It detects any variation in the environment, for instance the change of the noise level that affects the selection of the audio modality.
- Fission process: represents the fission rules/algorithm necessary to realise the fission.
- Subtasks-Modality (ies) Association: using patterns, this module allows us to select scenarios occurred previously and stored in the knowledge base.
- Modality selection: This module allows selection of the appropriate modalities available for each sub-task.
- Knowledge base: describes the environment, the modality patterns and the patterns of scenarios that occurred previously. It is described briefly in C.2.

A detailed architecture of the system is shown in Fig 5. It contains six main modules: "Parser", "Fission", "Grammatical Analysis", "Feedback", "Synchronization" and " Pattern of Modality Selection."

These modules communicate together using XML files (Wyke, Rehman et Leupen, 2002) to exchange information. They are loaded onto a computer, robot, or any device that can communicate via the Internet, social networking, etc.

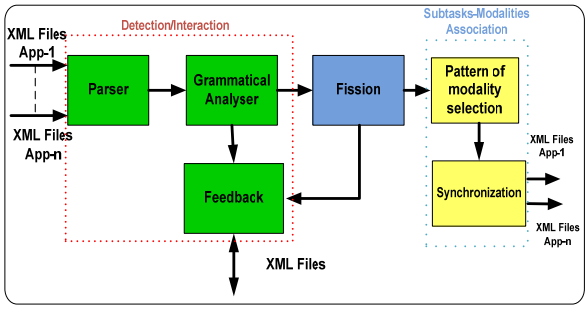


Fig 5: Framework of the multimodal fission.

The system is composed of the following elements:

- **Parser:** this module has as input XML files from different applications (for instance a robot control system or GPS system). Its role is to extract important information from the XML file (modality available, context, command... etc.) usable by the fission.
- **Fission:** based on the parameters of each modality and taking into account the rules of fission, this module determines whether the fission is possible and presents in the output the sub-tasks of each command.
- **Grammatical Analyser:** this module aims to check if the command is grammatically correct.
- **Feedback:** this module corrects the errors made by the user or by the recognition module by sending a feedback to the user.
- **Synchronization:** this module is designed to synchronize between the outputs for each modality.
- **Pattern of Modality Selection:** selects the appropriate modalities for each sub-task.

Vol. 4, No. 1 Jan 2013 ISSN 2079-8407
Journal of Emerging Trends in Computing and Information Sciences

http://www.cisjournal.org

b. Modality Selection and interaction context

Before we select the appropriate modality (ies) for every sub-task, we need to select the available modality using the interaction context.

In multimodal systems, context plays an important role to select the appropriate modalities. The selection module is called "interaction context". This module is shown in

Fig 6.

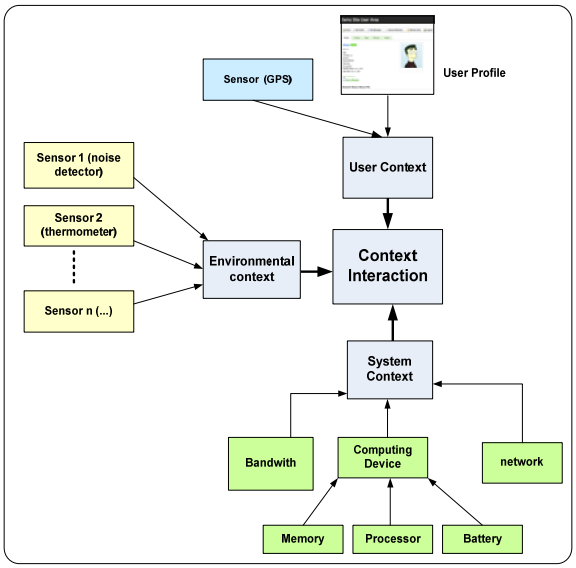


Fig 6: Interaction context module (Zaguia et al., 2010a).

This module is composed of three sub-modules:

- **User context**

This module detects the location and status of the user. It defines the user's ability to use certain modalities. For example, the system disables the display modality if it detects that the user is usually impaired and disable the audio modality if it detects that the user is in a library.

- **Environment context**

It describes the state of the environment such as determination of the noise level. It is understood that the use of the audio method (entered or left) is affected by this information. If the noise level is high, the audio modality will be disabled.

- **System context**

The capacity and the type of the system that we use, are factors that determine or limit the modalities that can be activated.

A modality is appropriate to a given instance of interaction context if it is found to be suitable to every parameter of the user context, the environmental context and the system context.

The suitable of specific output modality is shown by the relationships given below extracted from Table 1 to Table 5. Symbols √ and × are used to denote suitability and non-suitability, respectively.

$$VO_{out} = (user \neq deaf) \wedge (location \neq at\ work)$$

$$M_{out} = (user \neq manually\ handicapped) \wedge (location \neq on\ the\ go) \wedge (computer \neq cellphone/PDA \vee computer \neq iPad)$$

$$VI_{out} = (user \neq visually\ impaired) \wedge (workplace \neq dark \vee workplace \neq very\ dark)$$

Table 1: User location and its suitability to modalities.

Modalities	At Home	At Work	On the go
Vocal Output (VO_{out})	√	×	√
Manual Output (M_{out})	√	√	×
Visual Output (VI_{out})	√	√	√

Table 2: User handicap/profile and its suitability to output modalities

Modalities	Regular User	Deaf	Mute	Manually Handicapped	Visually Impaired
Vocal Output (VO_{out})	√	×	√	√	√
Manual Output (M_{out})	√	√	√	×	√
Visual Output (VI_{out})	√	√	√	√	×

Table 3: Noise level and its suitability to output modalities

Modalities	Quiet	Noisy
Vocal Output (VO_{out})	√	×
Manual Output (M_{out})	√	√
Visual Output (VI_{out})	√	√

Table 4: Brightness or darkness of the workplace and how it affects the selection of appropriate output modalities.

Modalities	Workplace Bright	Workplace Dark	Workplace Very Dark
Vocal Output (VO_{out})	√	√	√
Manual Output (M_{out})	√	×	×
Visual Output (VI_{out})	√	√	√

Table 5: The type of computing device and how it affects the selection of appropriate output modalities.

Modalities	PC/Laptop	Ipad	Cellphone/PDA
Vocal Output (VO_{out})	√	√	√
Manual Output (M_{out})	√	×	×
Visual Output (VI_{out})	√	√	√

In this work, the proposed system detects available media devices and produces result in which appropriate modalities are noted. An example is shown in Fig 7.

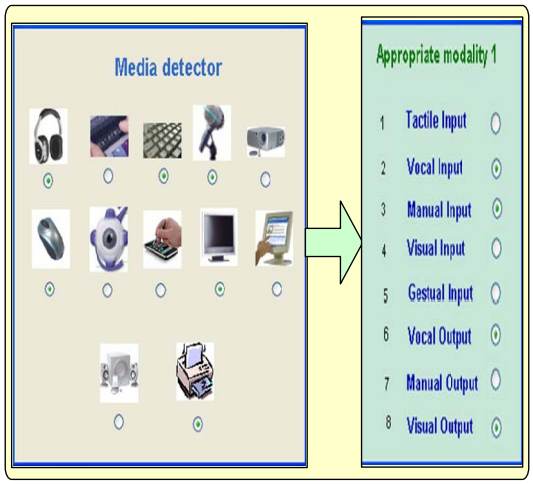


Fig 7: System detection of appropriate modalities based on available media devices.

The system detects the values of all related interaction context parameters and accordingly selects the appropriate modalities. A sample of such scheme is shown in Fig 8.

Fig 8: System detection of appropriate modalities based on the instance of interaction context.

Fig 9 illustrates how the optimal modalities are established – that is, finding the intersection between appropriate modality 1 (a set of modalities) and appropriate modality 2 (a set of modalities). In the cited case, the optimal modalities are vocal input, visual and vocal output.

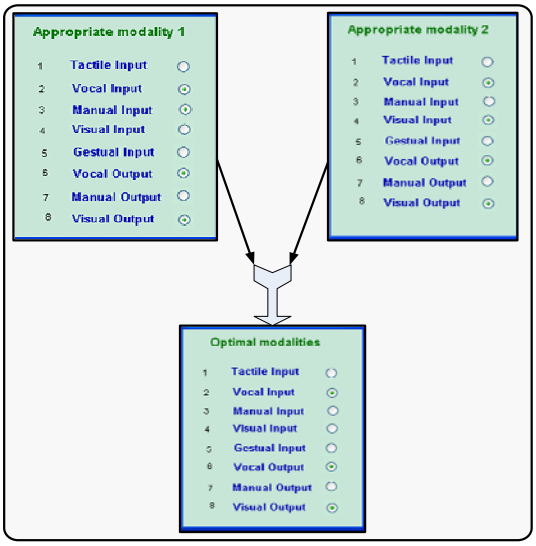


Fig 9: Optimal modalities – the results of the intersection between the set of appropriate modality 1 and the set of appropriate modality 2.

For more details concerning interaction context, the reader can refer to (Zaguia et al., 2010a).

c. Multimodal Fission

This module is the crucial component of our architecture. In this section, we describe our fission module with focus on the use of patterns as a solution to the described problems of fission systems find in literature.

We start by defining the pattern found in literatures and then we present how we use it to resolve our problem.

c.1 PATTERN

In this section, we define the pattern as found in the literature and we show how we adapt it in our case.

Alexander (Alexander, Ishikawa et Silverstein, 1977) has defined the pattern as a problem that often occurs in an environment. For Grone(Grone, 2006), "patterns help transporting knowledge and provide common names for solutions".

Generally patterns are defined with two parts, namely, problem and solution. Therefore, we must define the problem and the solution so that we can talk about patterns.

Vol. 4, No. 1 Jan 2013 ISSN 2079-8407
Journal of Emerging Trends in Computing and Information Sciences

http://www.cisjournal.org

c.1.1 Subtasks-Modalities Association

In this section, we present the use of pattern to associate the adequate modality (ies) to subtask.

Patterns are predefined models that describe a modality (ies) selection. In our work, a modality pattern is composed of: a) Problem composed of the components: Application, Parameter, Priority, Combination, Scenario and Service and b) Solution composed of the chosen modality as shown inFig 10.

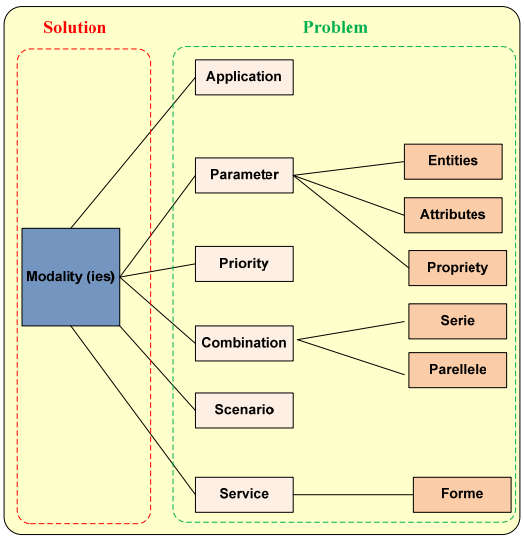


Fig 10: Pattern of modality (ies) selection.

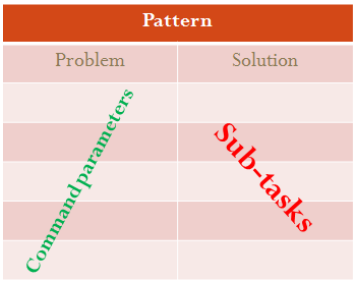
Our goal is to determine the best modality (ies) suited to a specific context (the solution: here is to choose a modality). We should define all the parameters that affect the choice of modality (problem: taking into account a number of parameters to choose this modality). So through these steps, we can model our patterns. Patterns will be stored in a knowledge base. The significance of each component in Fig 10is as follows:

- Application: Each modality is connected to a given application.
- Parameters: parameters of the pattern must contain elements that suit for a specific modality (entities, attributes, properties). These elements ensure the equivalence between data of a modality and the pattern itself (start time, end time, etc.).
- Priority: each modality has a priority depending on the application.
- Combination: modalities can be combined either in serial or in parallel with other modalities
- Scenario: each modality is connected to a scenario or sub-task.
- Service: each modality offers a service that can change shape. For example, if a modality offers the service display, the system increases or decreases the brightness (shape) depending on the level of light in the room.

c.1.2 Pattern of Sub-tasks Selection

For the pattern of sub-tasks, selection is composed of command parameters (problem) and the solution will be the sub-tasks (Table 6). Our patterns will be stored in a knowledge base.

Table 6: Pattern of sub-tasks selection.



Our goal is to select the adequate subtasks for each command. The command parameters (problem) are the words that compose the command. For example if the command is “Bring me the cup” the parameters of this command are “Bring-me-Cup” (problem) and the solution is {move to the cup-take the cup-move to the position (me)-depose the cup}

c.2. Fission Algorithm / Fission Rules

In general, the fission rule is simple: if a complex command (CC) is presented, then a set of sub-tasks with the modalities (and its parameters) are suitable deducted.

Multimodal fission can be represented by the function:

$$f : F \rightarrow ExK$$

$$\forall cc \in F, \exists STi \in K and MOj \in E, f(STi, MOj) = cc$$

with $i \in [1..n]$ and $j \in [1..m]$

$$f : CC = \sum_{i=1}^{n} ST_i \left(\left\{ \bigcup_{j=1}^{k} MO_j \right\}, \left\{ \bigcap_{j=1}^{l} MO_j \right\} \right) \quad (1)$$

with:

ST = sub-task.
MO = output modality.
CC= complex command.

l and k are different from m and n because it depends on the sub-tasks. For example, for some sub-task we will use just two terms even if we have three modalities available.

In equation (1), the symbol $\bigcup$ indicates that we can use either one or several modalities to present a sub-task. For example, if we present a text to the user, we use audio or display. The symbol $\bigcap$ indicates that we use the available modalities together to present a sub-task.

Vol. 4, No. 1 Jan 2013 ISSN 2079-8407
Journal of Emerging Trends in Computing and Information Sciences

http://www.cisjournal.org

Stages of the fission process are described in Fig 11, which shows the proposed fission algorithm. We assume that each XML file contains data for output modalities, their corresponding parameters and the complex command.

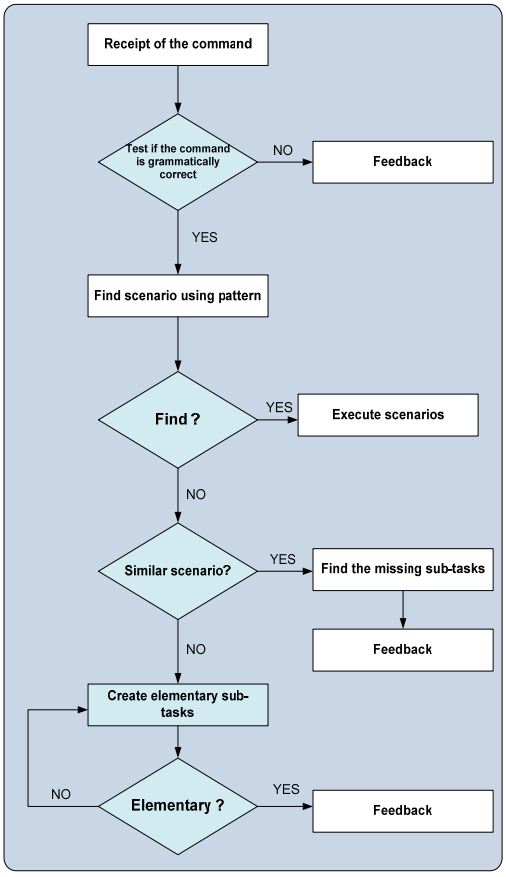


Fig 11: Fission algorithm.

When the system receives an XML file, as shown in

Fig 12, it extracts the command and it checks if it is complete by checking grammar rules. We defined many grammar rules among these rules for instance:

Fig 13). From

Fig 13, the subtask related to liquid (milk) is “add milk” so this subtask is added to subtasks of “prepare coffee" and then request feedback from the user.

- AFMO →MO for example “drop the cup”
- AFP→P→MO for example “ give me the pen”
  with AFMO = action for movable object, MO = movable object and P= person

If this command is incomplete, the system sends a feedback to the user to correct the error. Otherwise, a search pattern (pattern of subtasks selection) scenarios in the knowledge base starts.

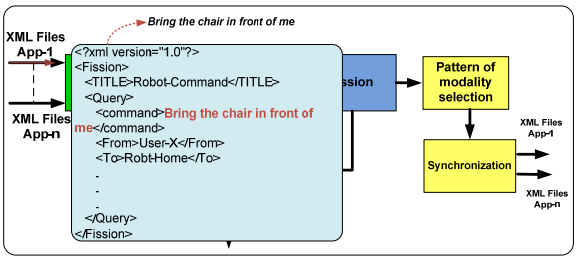


Fig 12: Extracting the command from XML file.

For example, if the command is "bring the chair in front of me," in this case, we look sub-tasks with sending a query (problem: bring object position) to the knowledge base.

If the scenarios found, the system performs sub-tasks. Otherwise, the system will search if there is a similar scenario. For instance, if in our knowledge base there is a scenario "prepare coffee" and the system receives a command "prepare coffee with milk" in this case, we have similar command so we take the result for "prepare coffee" and the system creates missing sub-tasks related to milk using our knowledge base(

In the case where the system cannot find a similar command, it creates elementary sub-tasks. The system asks the user to confirm if the result is correct.

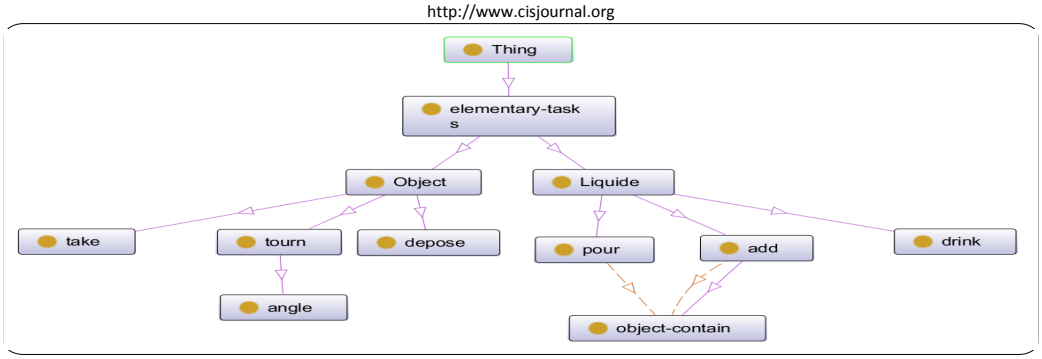


Fig 13: Knowledge Base of elementary tasks.

Suppose the system receives a command "Bring the chair in front of me" with the current context = living room. We can define a query(problem: bring object position), which will contain the necessary elements to search in the knowledge base if the scenario has already occurred.

With Action = Bring, Object = the chair and position= in front of me. Suppose we have patterns registered in our knowledge base as shown in

Fig 14.

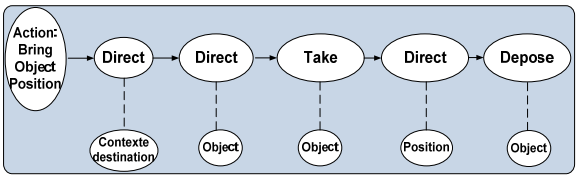


Fig 14: Example of pattern.

The search result is shown in

Table 7. In

Fig 14, "destination context" is the same as the "current context", adding "direct (destination context)" will allow us to manage orders for other different contexts position. For example, if the command was "Bring me a spoon" and the current context = living room. Since the spoon is in the kitchen so the destination context = kitchen.

Table 7: Pattern with the solution.

Pattern	
Problem	Solution
Bring	Direct to context-Destination
Object	Direct to Pos-object
Position	Take object
	Direct to position
	Depose object

5. SIMULATION

A handicap user sitting in the living room, uses his mobile device to send a message "bring me that" and point with his hand to a cup, to a robot located behind him. An XML file is created and sent to the fusion module as shown in Fig 15. The file contains the command, the available modalities and the context. After the process of fusion, this module sends the complex command to the fission module as shown in Fig 15: "bring me the cup" with the coordinate of the cup and the user.

Fig 15: Example of scenario.

To validate the stages of fission process, we use the collared Petri Net (CPN) (Jensen, 1987). The CPN is more advantageous than the ordinary Petri Net.

There are many simulation tools, often free, and developed in the context of thesis or scientific research. We used CPN Tools(Zhu, Tong et Cheng, 2011), one of the most used software, simulation of the high-level Petri network.

A CPN is a graphical structure linked to computer language system. The CPN is a Petri net in which the tokens are collared. The colour is an information attached to the token. This information allows distinguish between tokens and it can be any type (integer, real, String, Boolean, list etc.). It is based on a set of conditions and expressions that permit to the tokens to change their colours (change the state of the place).

Vol. 4, No. 1 Jan 2013 ISSN 2079-8407
Journal of Emerging Trends in Computing and Information Sciences

http://www.cisjournal.org

Fig 18shows the diagram of the example "Bring me the cup". This diagram shows the stages that the CPN validates to prove that system works properly.

Fig 19 shows the framework of the multimodal fission with CPN. In the input, we have the command "bring me the cup" as a string token. In the following diagrams, we present the validation of the fission module.

Fig 20 shows the parser module. We extract every word from the command to perform grammatical analysis.

Fig 21 to Fig 23show CPN for the communication between T_Grammar and Knowledge Base to validate if a command is correct and the creation of query for the search of sub-tasks. In our case the answer is "AFP person Mo" with:

- AFP: action for person= Bring
- person= me
- Mo: movable object = cup

Fig 25).

T_Fission module uses query that contains elements as shown in Table 8 to search the adequate pattern-subtasks in the knowledge base (

Table 8: Query for "AFM person Mo" with AFM = bring, person=me and Mo= cup.

Problem
AFM
Person
Mo

Fig 24presents the query "AFM person Mo" sent to the knowledge base to find the matching pattern.
Figure 25 presents the solution of the problem "AFM person Mo" : ["1-move to the object",
"2-take the object","3-move to the position", "4-depose the object"];

Figure 26 shows the final subtasks to be executed by the system:

a. move to the object
b. take the object
c. move to the position
d. depose the object

Fig 16: Declaration of variable.

Fig 16 and Figure 17 show the declarations of variables and functions used with CPN for this scenario.

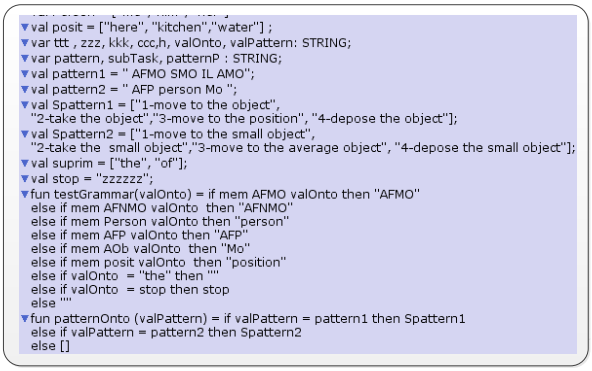


Fig 17: Declaration of variables.

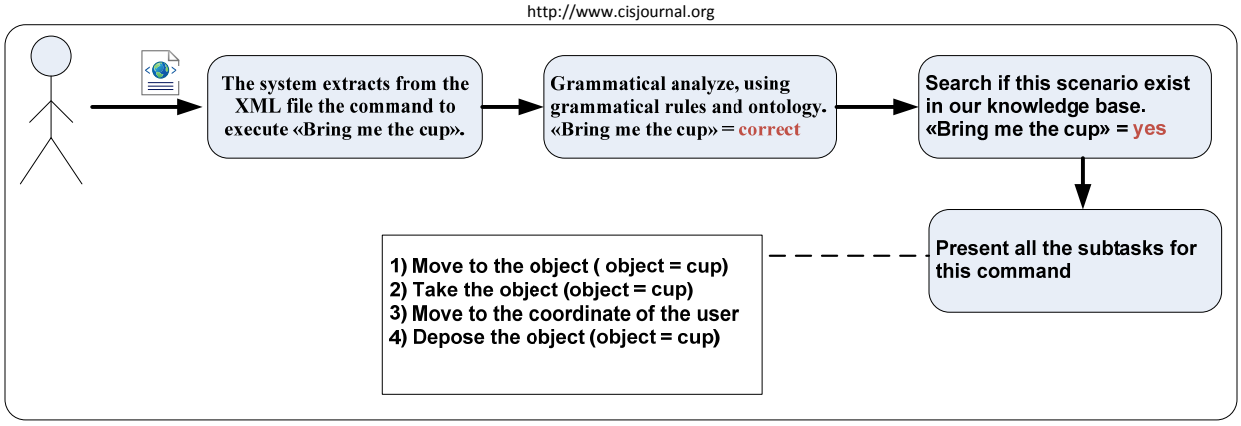


Fig 18: Example of scenario.

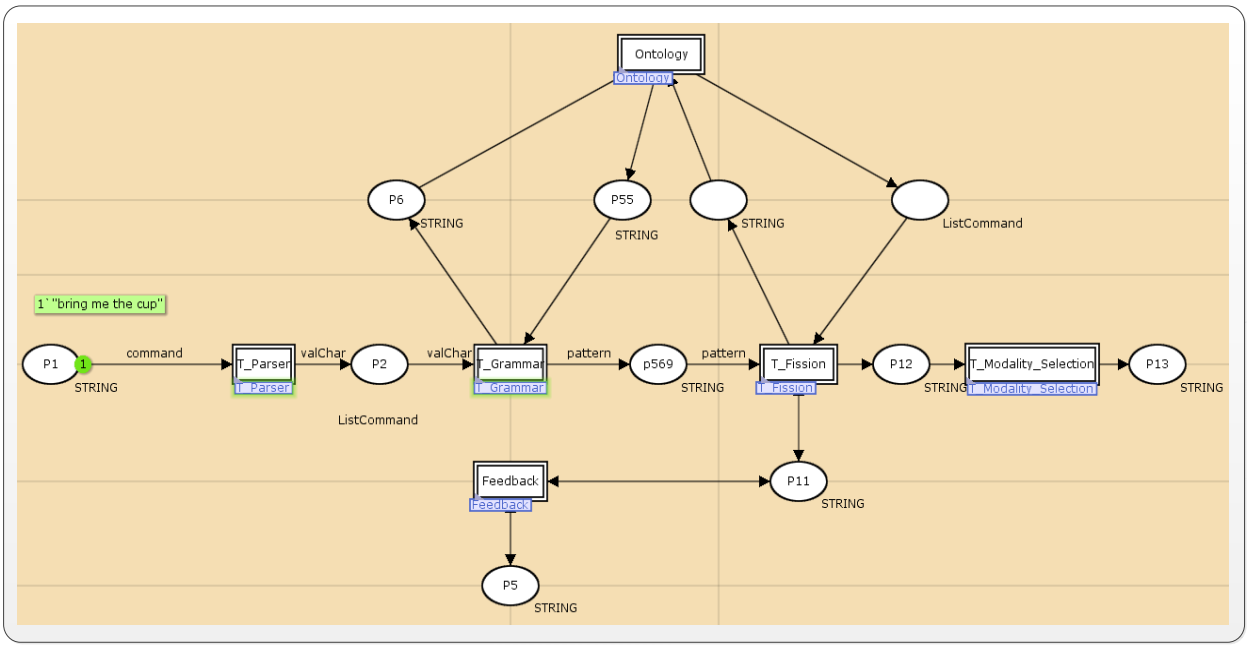


Fig 19: Framework of the multimodal fission with CPN.

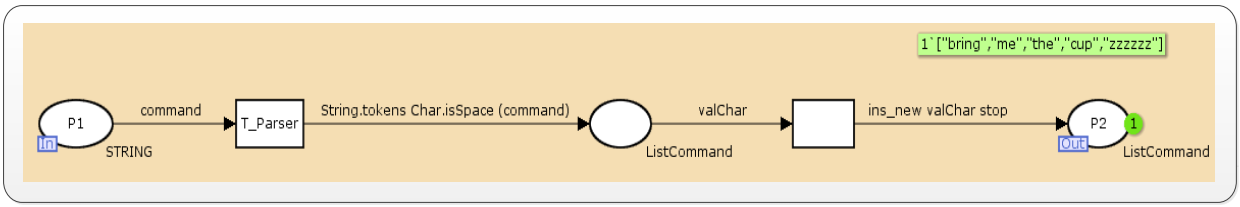


Fig 20: Colored Petri Net showing the operation of parser module.

Vol. 4, No. 1 Jan 2013 ISSN 2079-8407
Journal of Emerging Trends in Computing and Information Sciences

http://www.cisjournal.org

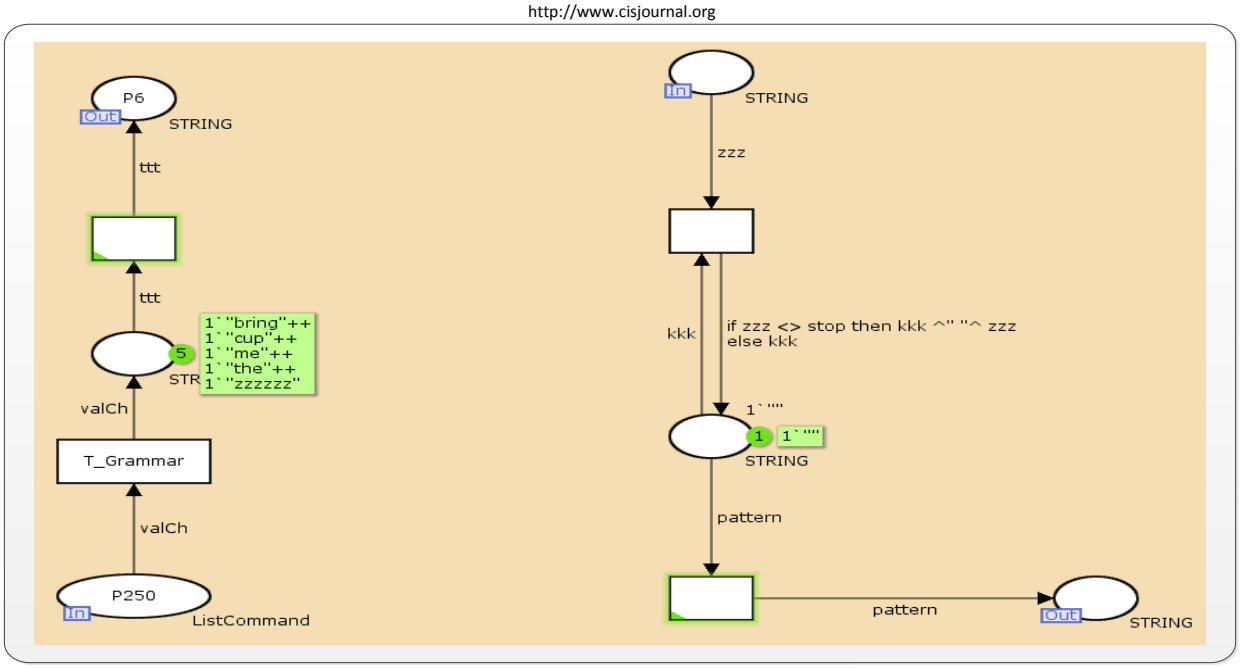


Fig 21: Colored Petri Net showing the operation of Grammar module.

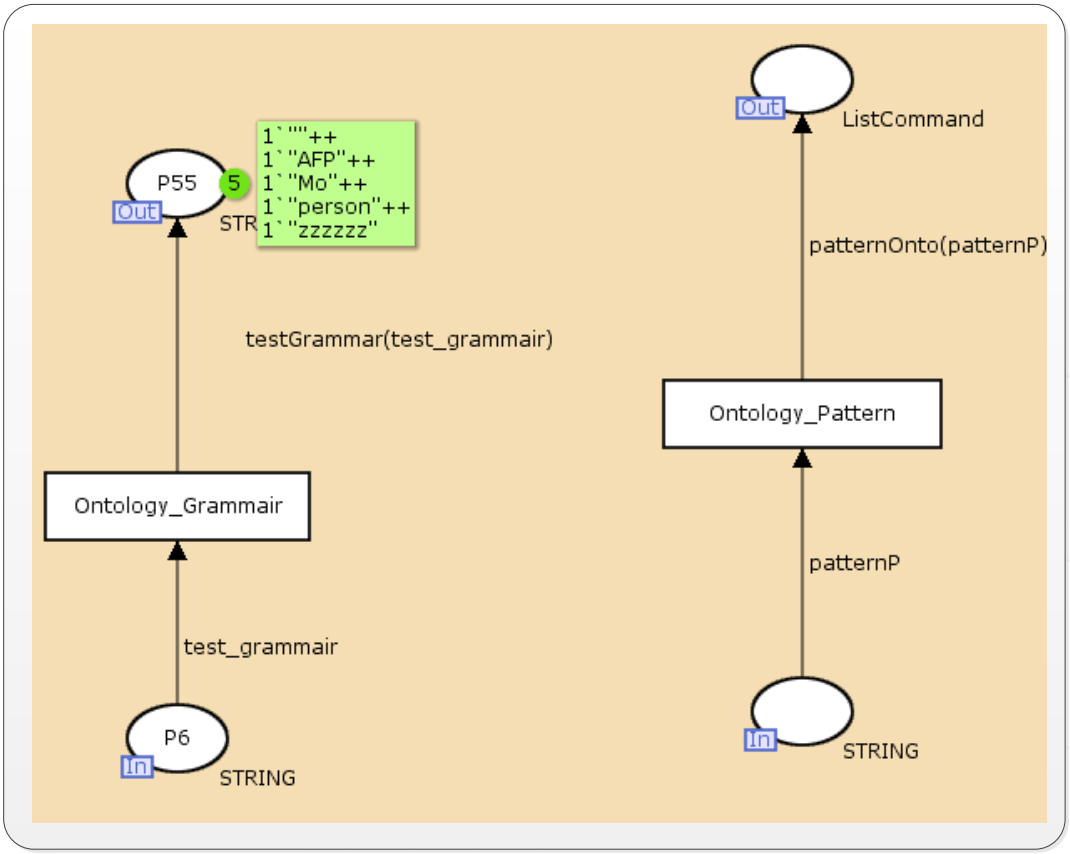


Fig 22: Colored Petri Net showing the operation of ontology concerning the grammar.

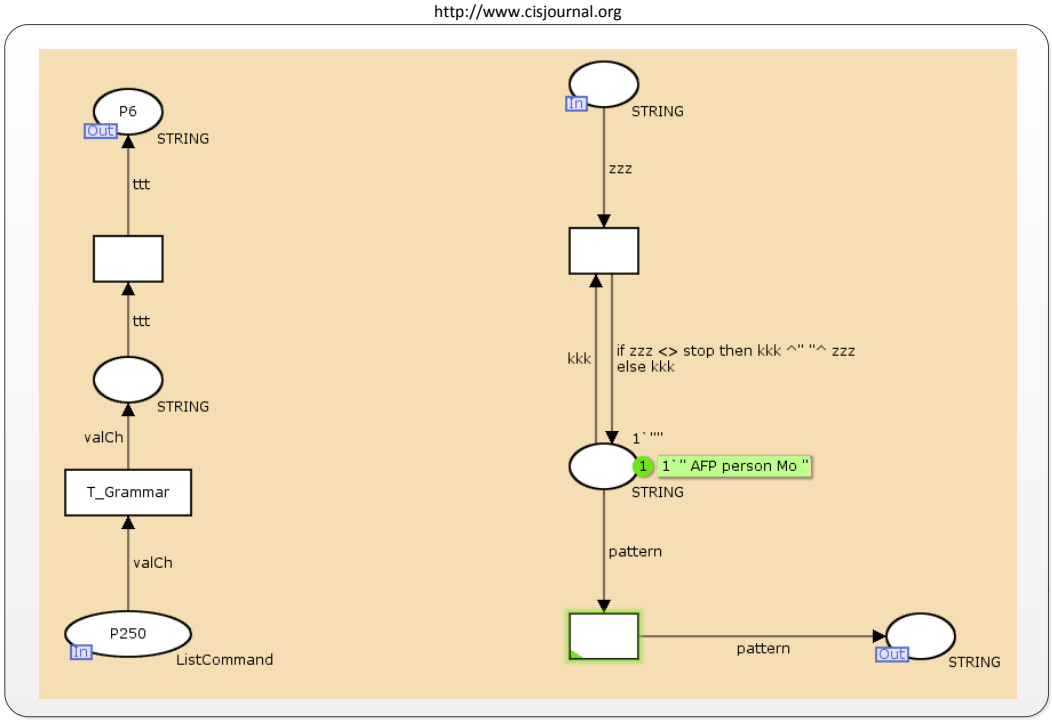


Fig 23: Colored Petri Net showing the creation of pattern to find sub-tasks.

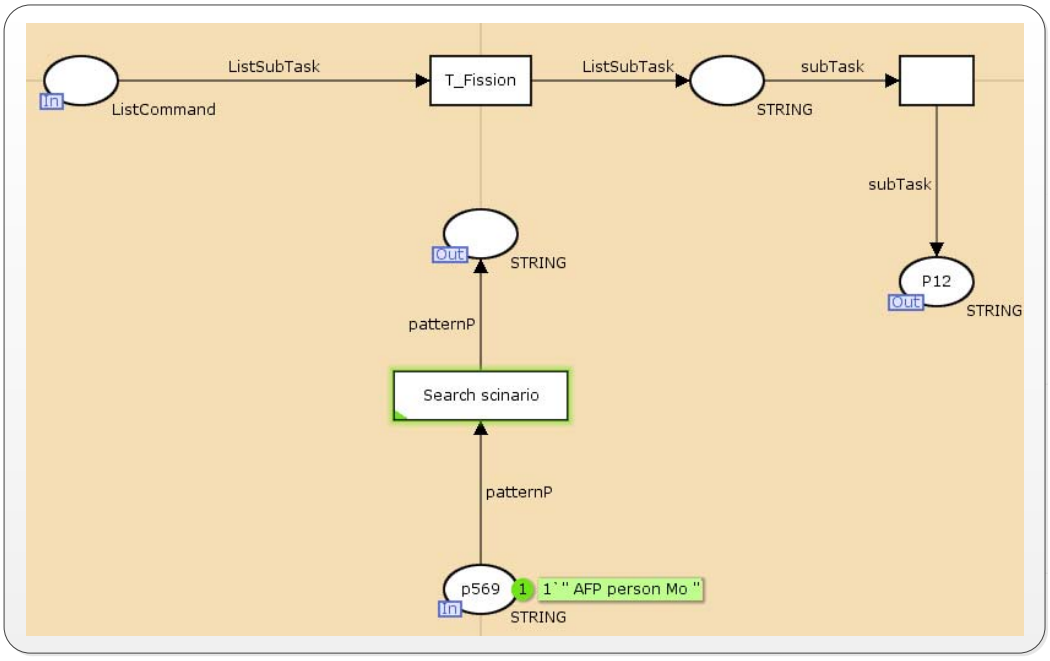


Fig 24: Colored Petri Net showing the search of sub-tasks using pattern.

Vol. 4, No. 1 Jan 2013 ISSN 2079-8407
Journal of Emerging Trends in Computing and Information Sciences

http://www.cisjournal.org

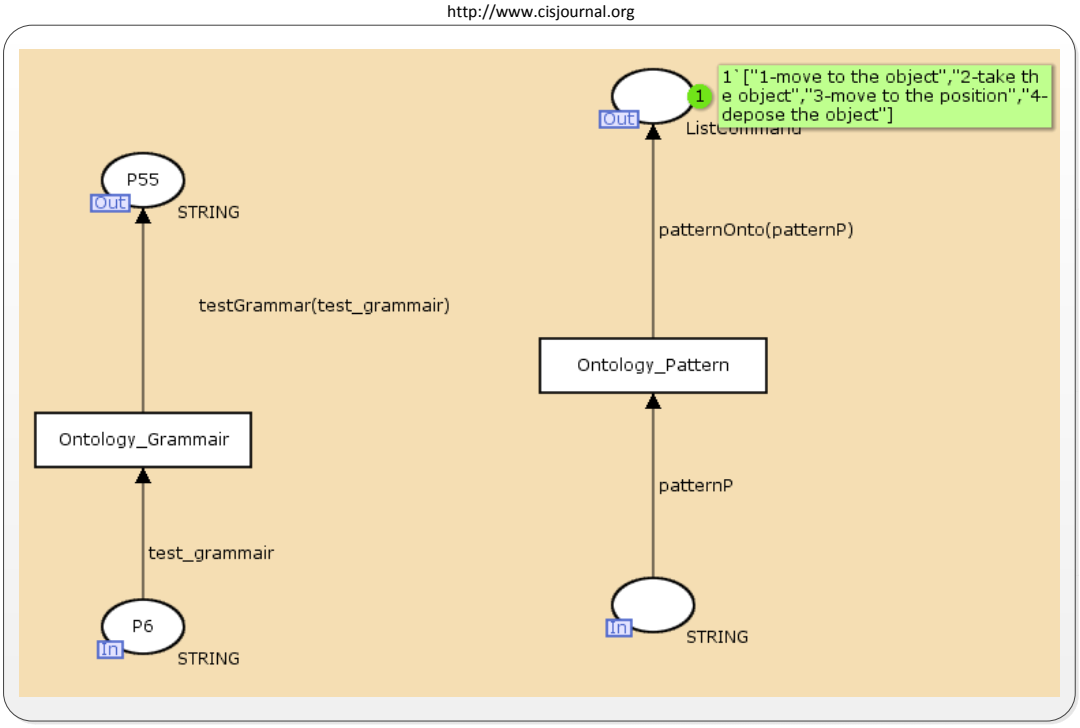


Fig 25: Colored Petri Net showing the sub-tasks found for the pattern "AFP person Mo"

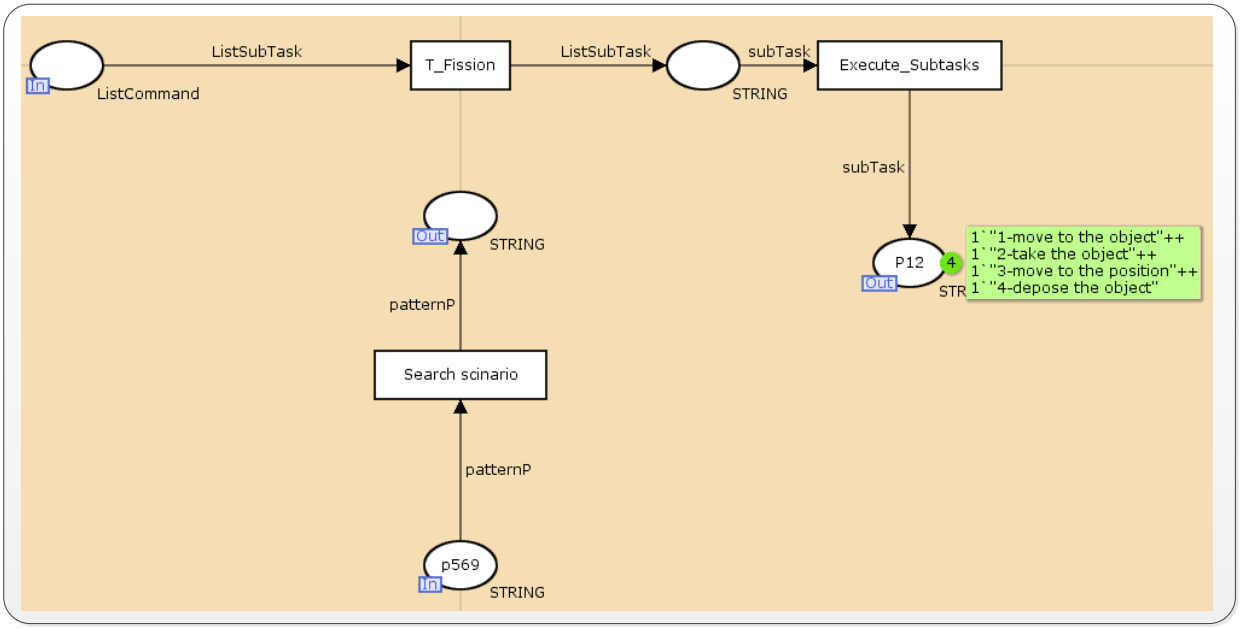


Fig 26: Colored Petri Net showing the execution of sub-tasks

6. CONCLUSION

In this article, we presented an architecture which is very useful in a multimodal system. We presented an effective algorithm for the fission process. We have shown the important role of the fission module. This module subdivides a complex command into elementary sub-tasks and presents them to the output modalities. The proposed solution facilitates the work of the fission module, using predefined patterns stored in a knowledge base or ontology. These patterns save time and facilitate the executions of the tasks. They can also be used by other researchers in their own work. A concrete example is illustrated in order to show the effectiveness of the contribution, using CPN tool.

This project has several scientific benefits. It will promote research in this area, contribute to the advancement of existing knowledge in the human-machine interaction domain generally and fission specifically, and provide researchers with a fission engine and a set of patterns that can be very useful. On the other hand, like any scientific research project, the essential goal remains humanity with making life easier for others especially for the elderly, disabled or sick persons. The created architecture can be deployed on robots, mobile devices, computers, etc.

ACKNOWLEDGEMENT

We wish to acknowledge the funds provided by the Natural Sciences and Engineering Council of Canada (NSERC) which support the financial needs in undertaking this research work.

REFERENCES

[1] Alexander, Christopher, Sara Ishikawa and Murray Silverstein (1216). 1977. A Pattern Language: Towns, Buildings, Construction. Center for Environmental Structure Series. Published August 25th 1977 by Oxford University Press, USA. 1261 pages

[2] Alm, Torbjorn, Jens Alfredson and Kjell Ohlsson. 2009. « Simulator-based human-machine interaction design ». International Journal of Vehicle Systems Modelling and Testing, vol. 4, n° 1/2, p. 1-16.

[3] Atrey, Pradeep, M. Hossain, Abdulmotaleb El Saddik and Mohan Kankanhalli. 2010. « Multimodal fusion for multimedia analysis: a survey ». Multimedia Systems, vol. 16, n° 6, p. 345-379.

[4] Beinhauer, Wolfgang, and Cornelia Hipp. 2009. « Using Acoustic Landscapes for the Evaluation of Multimodal Mobile Applications ». In Human-Computer Interaction. Novel Interaction Methods and Techniques. San Diego, CA, USA.

[5] Benoit, Alexandre, Laurent Bonnaud, Alice Caplier, Phillipe Ngo, Lionel Lawson, Daniela G. Trevisan, Vjekoslav Levacic, Céline Mancas and Guillaume Chanel. 2009. « Multimodal focus attention and stress detection and feedback in an augmented driver simulator ». Personal and Ubiquitous Computing, vol. 13, n° 1.

[6] Bolt, R. 1980. « Put-that-there ». Voice and gesture at the graphics interface ACMSIGGRAPH Computer Graphics, vol. 14, n° 3, p. 262-270.

[7] Caschera, Maria Chiara, Alessia D'Andrea, Arianna D'Ulizia, Fernando Ferri, Patrizia Grifoni and Tiziana Guzzo. 2009. « ME: Multimodal Environment Based on Web Services Architecture ». In OTM 2009 Workshops. (Vilamoura, Portugal), p. 514-512. Springer.

[8] Costa, David, and Carlos Duarte. 2011. « Adapting Multimodal Fission to User's Abilities Universal Access in Human-Computer Interaction. Design for All and eInclusion ». In the 6th international on Universal access in human-computer interaction: design for all and eInclusion. Orlando, FL.

[9] Ertl, Dominik, Jürgen Falb and Hermann Kaindl. 2010. « Semi-Automatically Configured Fission For Multimodal User Interfaces ». In Third International Conference on Advances in Computer-Human Interactions. (Saint Maarten, Netherlands, Antilles).

[10] Foster, Mary Ellen. 2002. State of the art review: Multimodal Fission. University of Edinburgh. COMIC project.

[11] Foster, Mary Ellen. 2005. « Interleaved Preparation and Output in the COMIC Fission Module ». In Software '05 Proceedings of the Workshop on Software (Stroudsburg, PA, USA).

[12] Grone, Bernhard. 2006. « Conceptual Patterns ». In ECBS '06 Proceedings of the 13th Annual IEEE International Symposium and Workshop on Engineering of Computer Based Systems.Washington, DC, USA.

[13] Jensen, Kurt. 1987. « Coloured Petri nets Petri Nets: Central Models and Their Properties ». Vol. 254, p. 248-299. Berlin (Heidelberg) : Springer.

[14] Landragin, Frédéric. 2007. « Physical, semantic and pragmatic levels for multimodal fusion and fission ». In Seventh International Workshop on Computational Semantics. Tilburg, The Netherlands.

[15] Meng, H., S. Oviatt, G. Potamianos and G. Rigoll. 2009. « Introduction to the Special Issue on Multimodal Processing in Speech-Based Interactions ». Audio, Speech, and Language Processing, IEEE Transactions on vol. 17, n° 3, p. 409 - 410

[16] Poller, Peter, and Valentin Tschernomas. 2006. « Multimodal Fission and Media Design ». In SmartKom: Foundations of Multimodal Dialogue Systems, sous la dir. de Wahlster, Wolfgang. Springer Berlin Heidelberg.

[17] Robert, Steele, Khankan Khaled and Dillon Tharam. 2005. « Mobile Web Services Discovery and Invocation Through Auto-Generation of

Abstract Multimodal Interface ». In ITCC 2005 International conference on Information Technology. Las Vegas, NV. Vol. 35-43. IEEE.

[18] Sears, Andrew, and Julie A. Jacko. 2007. The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, Second Edition. CRC Press, 1,384 p.

[19] Wehbi, Ahmad, Manolo Hina, Atef Zaguia, Amar Ramdane-Cherif and Chakib Tadj. 2011.« Patterns Architecture for Fusion Engines ». 9th International Conference on Smart Homes and Health Telematics, Montréal, Quebec, Canada, June 2011.

[20] Wyke, Allen, Sultan Rehman and Brad Leupen. 2002. Manuel de Référence XML. Vivendi Universal Publishing Services, 706 p.

[21] Zaguia, Atef, Manolo Dulva Hina, Chakib Tadjand Amar Ramdane-Cherif. 2010a. « Interaction Context-Aware Modalities and Multimodal Fusion for Accessing Web Services». Ubiquitous Computing and Communication Journal, vol. 5, n° 4.

[22] Zaguia, Atef, Manolo Dulva Hina, Chakib Tadj and Amar Ramdane-Cherif. 2010b. « Using Multimodal Fusion in Accessing Web Services ». Journal of Emerging Trends in Computing and Information Sciences, vol. 1, n° 2, p. 121-138.

[23] Zhu, Lulu, Weiqin Tong and Bin Cheng. 2011. « CPN Tools' Application in Verification of Parallel Programs Information Computing and Applications ». In Information Computing and Applications. Vol. 105, p. 137-143. Springer Berlin(Heidelberg).