DISA 2018

1st World Symposium on Digital Intelligence for Systems and Machines

PROCEEDINGS

IEEE

August 23-25, 2018
Košice, Slovakia

DISA 2018

IEEE World Symposium on Digital Intelligence for Systems and Machines

PROCEEDINGS

August 23-25, 2018
Košice, Slovakia

Website: http://www.disa2018.org/

ISBN: 978-1-5386-5101-8
IEEE Catalog number: CFP18P13-USB
Proceedings editor: Ján Paralič, Peter Sinčák, Laszlo Kovacs, XenFuWang

TABLE OF CONTENTS

Invited papers

Various Aspects of Digital Intelligence

Intelligent Data Analysis with Special Focus on Analysis of Texts and Discursion Content in Social Networks (Special Session)

Intelligent Socio-Cyber-Physical Systems and Internet of Things in Interconnected World (Special Session)

VoMIS – the VoiceXML-based multimodal interactive system for NAO robot

Stanislav Ondáš,MatúšPleva, Radovan Krištan, Rastislav Husovský, JozefJuhár
*Department of Electronics and MultimediaCommunications, Faculty of ElectricalEngineering and Informatics, TechnicalUniversity of Košice,
{stanislav.ondas, matus.pleva,jozef.juhar}@tuke.sk

***Abstract*— The proposed paper brings a description of the VoiceXML-based multimodal interactive system (VoMIS) for NAO humanoid robot. The designed system enables the multimodal interaction with the user in such manner that it takes a speech input from the user and it answers by a combination of synthetic speech and gestures. The core of the system is an external dialogue manager VoiceON, which interprets VoiceXML language.VoiceXML was originally designed for unimodal systems, but thanks to its advantages we decided to extent it to manage multimodal interactions. Our work illustrates how VoiceXML can be easily extended to manage also multimodal interaction, mainly using <prompt> element. Designed changes enables to control movements and gestures of the robot.**

Keywords – multimodality, human-robot interaction, VoiceXML, dialog

I. Introduction

The importance of speech technologies grows thanks to new types of devices with new types of interfaces as are smartphones, smart watch, virtual reality (VR) devices (glasses) or other household appliances, that are designed in the concept of internet of things.

Humanoid robotics can be identified as the next very important area, where speech technologies are located. The result of undeniable effort to develop artificial human is several humanoid robots (e.g. Honda Asimo, Aldebaran NAO, Pepper) with different capabilities and with a different degree of human-like appearance [1]. With an improved appearance of humanoid robots, people tend to expect a human-like behavior, including speech production [2] and hearing capabilities.

The communication between people has a multimodal character. It means that information is usually shared through several modalities. To make human-machine interaction more human-like, multimodal interaction need to be implemented into HMI.

In case of humanoid robots, gestures together with speech can be considered as very appropriate channels for information exchange. Combination of speech and gestures supports natural and effective human-robot interaction. Moreover, gestures can convey so called “backchannel signals”, which can significantly contribute to smooth interaction[3], [4].

To make the process of preparing dialogue interactions with NAO robot more effective and simpler, we designed and developed the multimodal interactive systemVoMIS (the VoiceXML-based multimodal interactive system).It is the asymmetric multimodal dialogue system, which allows users to interact with robot by voice and robot uses speech and gestures to answer [5].

The external dialogue manager VoiceON [6] is a core of the system. It serves for managinghuman-robot interaction based on interpretation of VoiceXMLscripts that can be extended with instructions for robot gestures and movements.

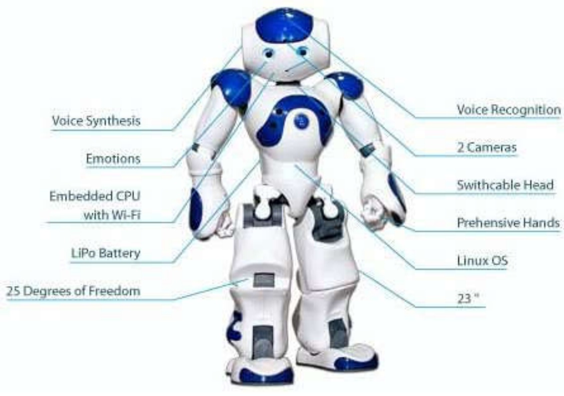

Figure 1. Humanoid robot NAO (Softbank Robotics)

The proposed work introduces VoMIS system, its components and shows how this system can be used with NAO robot. It describes an extension of the VoiceXML language to support generation of multimodal output through NAO. Described system enables to control robot’s gestures and movements. The VoiceXML language has been selected thanks to its advantages, mostly good readability and maintainability of the created speech-based interactions.

The paper is organized as follows: The second section describes designed multimodal dialogue system.The next section focuses on multimodal output generation and movement control. Section IV. brings more details about VoiceXML application that serves for managing interaction with the robot. Section V. deals with a comparison of the VoMIS system with the built-in dialogue system.

II. VoMIS – the VoiceXML-based Multimodal Interactive System

The basic architecture of the VoMIS system is shown in Fig. 2. VoMIS is the asymmetric multimodal dialogue system. Actually, it supports input in the form of user’s speech and it enables to provide output in a form of speech, gestures and movements combination. VoMIS can be extended to symmetric system, by adding other input communication modalities (e.g. user’s gestures recognition). From the system's point of view, NAO robot can be seen as an interface between VoMIS and user, because it enables to take user's input and it brings the system's output back to the user.

Humanoid robot NAO (see Fig. 2) is an autonomous programable robot originally developed by Aldebaran Robotics company (actually Softbank Robotics). NAO can be considered as a great tool to prepare a multimodal dialogue system due to its support of vision, hearing, gesture production and body language. Moreover, an autonomous mode of the robot supports a human-like behavior.

NAO operation system and development kit enables to prepare multimodal interactions also with built-in modules. In addition to ASR and TTS systems that supports 19 languages, NAO contains also a dialogue module, which enables to write spoken interactions. Built-in dialogue module is called ALDialog. The ALDialog module allows to endow robot with conversational skills by using a list of "rules" written and categorized in an appropriate way [7]. Despite the fact, that ALDialog module enables very natural conversations, it lacks a stronger flow management, which cannot by written directly into the rules and need to be written into the source code. Moreover, writing dialogue interactions for NAO cannot be considered as simple and effective.

These facts lead as to start think about our own dialogue unit, which can offer more simpler development of spoken dialogue with the user. Moreover, we prefer an external unit, to make possible to control several robots in the same time and to share knowledge between them through cloud.

In the proposed dialogue system, we adopt our earlier developed VoiceXML-based dialogue manager VoiceON[6], which we modified to enable multimodal interaction.

A. VoMIS Architecture

VoMIS architecture is distributed, what means that one part of the system run inside the robot but dialogue manager run on external server. In case of using other than built-in ASR (Automatic Speech Recognition) and TTS (Text-to-Speech) engines, they can also run outside the robot.

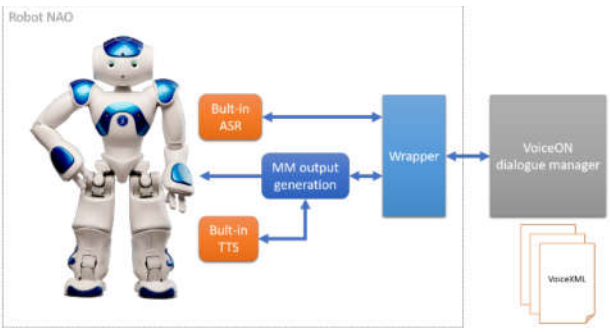

Figure 2. The architecture of the multimodal dialogue system for NAO robot.

The communication between NAO functionalities and the dialogue manager is transformed in wrapper module. The system uses built-in ASR module and TTS modules, which do not support Slovak language. To enable communication in Slovak language, we use Czech localization of these services inside NAO.

To produce multimodal output of the robot the new module was designed – MultiModal output generation module, which serves for generating output behavior of the system in a form of speech, gestures and movements.

B. Dialogue manager VoiceON

VoiceON dialogue manager is an interpreter of VoiceXML 1.0 language, which was designed in range of national project IRKR (see [8] or [9]). It has client-server architecture, where the client side can be easily rewritten for new applications.

The mechanism of dialogue management lies in interpretation ofVoiceXML scripts, which define the content and flow of the interaction. VoiceXML language enables to prepare system-directed interactions and interactions with mixed initiative.

VoiceON has been adopted to control dialog in VoMIS dialogue system. There are two main advantages of using VoiceON. Manager is a server-client system, what enables to let the main part of manager on the server side. This solution enables dynamically change the dialogue content by generation of VoiceXMLscript on the server-side.

The structure of VoiceON manager is shown in Fig. 3.

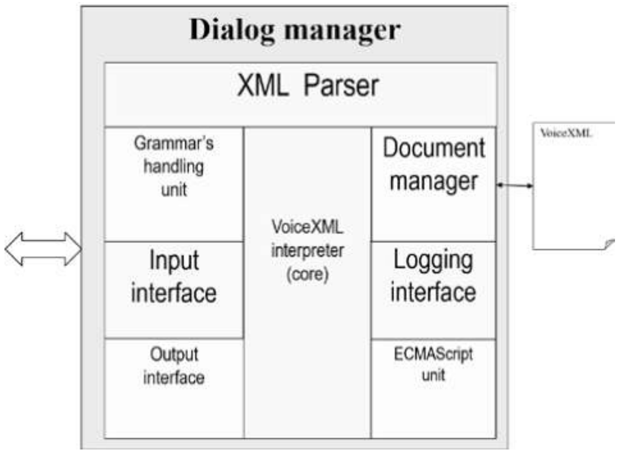

Figure 3. VoiceXML-based dialogue manager VoiceON

To integrate VoiceON into the VoMIS system, the new client of the manager (Wrapper) has been created to map messages of the manager into the instructions for the robot and vice versa. Wrapper run inside the robot.

III. Multimodal Output Generation and Movement Control

The multimodal output generation module (MM output generation) serves for generation of the output behavior in a form of speech and gestures and robot movements. The module is integrated directly inside the robot and is triggered from the wrapper of the dialogue manager. It analyses the content of the <prompt> element to decide about the appropriate combination of speech and gestures or about triggering of the particular movements.

A. Generation of speech and gestures

NAO robot can generate a multimodal output in a form of speech and gestures. Unfortunately, this functionality is supported only for English language. Therefore, we have started to develop a new multimodal output generation module (see Fig. 4) to generate a meaningful output that consists of speech and gestures.

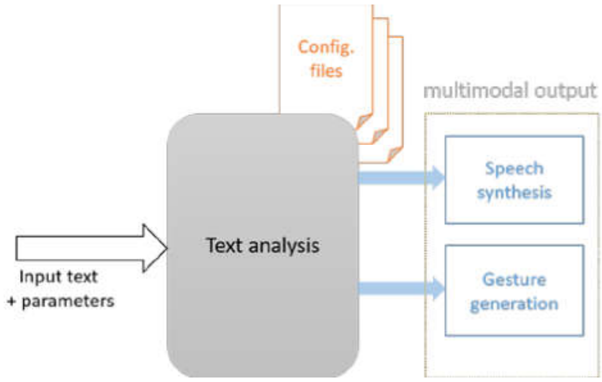

Figure 4. Multimodal output generation module

Designed module takes an input text and perform the analysis, which split the text into simple sentences and looks for keywords. Then manually written rules, stored in configuration files, are applied to add appropriate gesture marks into the output stream. The module can work in three different modes:

- **Speech only**. Output is generated without gestures.
- **RandOff**. Speech and gestures are generated. Gestures are generated as the output of the text processing. Robot random mode is switched off.
- **RandOn**. Speech and gestures are generated. Gestures and movements are produces both as the output of the text processing and as a result of Robot random behaviour.

B. *Movement control*

ALMotion and ALRobotPosture modules of the robot were used. Module ALMotion control walking of the robot using walkTo(x,y,z) function, where x, y and z are parameters for walking. Module ALRobotPosture offers methods for controlling the basic body posture of the robot.

To control movements, wrapper of the VoiceON dialogue manager together with output generation module were modified. Movements are triggered from AddMovement module.

On the side of the VoiceXML script, we implemented the movement control into the <prompt>VoiceXML element by overloading its content. In case of movements, the syntax of the <prompt> content is as follows:

```
#action_direction_distance
```

This simple syntax enables to deliver required action with two parameters – direction and distance/speed. Prompt element with movement command may look like:

```
<prompt>#move_forward_3 </prompt>
```

In case that movement command occurs inside the prompt, AddMovement module is activated, which decodes the meaning of the command and transform it into ALMovement or ALRobotPosture function call.

AddMovement module is written in Python and run inside the robot (Addmovement.py).

The whole VoiceXML script with the movement command can be seen in Fig. 7.

VoMIS system supports following movement commands:

- `move_forward, move_back, move_left, move_right`
- `move_turnleft, move_turnright,`
- `move_sit, move_stand, move_crouch`
- `move_lyingBack, move_lyingBelly`
- `move_standZero,move_standInit, move_sitRelax`
- `move_Hello`

The last parameter of the movement command can be reserved for distance or speed. In case of ALRobotPosture commands (Stand, Sit), this parameter means speed of posture realization.

The next figure illustrates implemented commands for movements.

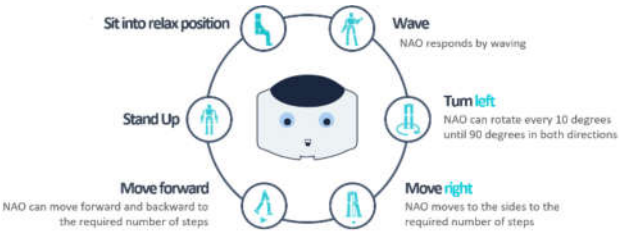

Figure 5. Movement commands

IV. VoiceXML Application

Figures 6 and 7 illustrate the simple VoiceXML script, which enables to use movement commands during the spoken interaction between human and NAO and a processing flow of the VoiceXML application.

Figure 6. Processing flow of the VoiceXML script

Interaction starts, when the dialogue manager starts to interpret the VoiceXML file. Interpretation begin with the first <form> element in the document order. After loading the speech grammar defined in <grammar> element, behavior specified in <prompt> element is performed, automatic speech recognition is activated, and NAO start to listen to the user's speech. Provided user's input is recognized by automatic speech recognition and then it is compared to active speech grammar. If user's speech matches rules in the grammar, such input can be interpreted by semantic markups included in the grammar. Then, the result of the semantic processing is delivered back to VoiceXML script, where it is processed in <filled> element.

Inside the <filled> element, appropriate prompts are generated, which contains instructions for movements or combination of speech and gestures, or for speech only output. The content of <prompt> element is processed by the multimodal generation module, which triggers actions using built-in robot's modules.

```
<?xml version="1.0" encoding="utf-8"?>
<vxml version="1.0" lang="slovak">

  <form id="command">
    <field name="mycom">
     <prompt bargein="false"> Čakám na povel. </prompt>
     <grammar src="commands.xml"
              type="application/grammar+xml">

     <filled>
       <if cond="mycom='forward1'">
         <prompt bargein="false"> #move_forward_1 </prompt>
         <goto next="done"/>
       <elseif cond="mycom='backward1'"/>
         <prompt bargein="false"> #move_backward_1 </prompt>
         <goto next="done"/>

       ...

       <elseif cond="mycom='moveright1'"/>
         <prompt bargein="false"> #move_right_1 </prompt>
         <goto next="done"/>
       <else/>
        <prompt bargein="false"> Tento povel nepoznám.</prompt>
        <clear namelist="mycom"/>
       </if>
     </filled>
    </field>
  </form>

  <form id="done">
    <block> <goto next="#command"/> </block>
  </form>
</vxml>
```

Figure 7. Example of the VoiceXML application with movement commands.

V. Comparison with Built-in Interactive System

NAO enables to create a multimodal interactive system exclusively with built-in components – modules. To build dialogues for such system, ALDialog module need to be used, which interprets scripts written in QiChat syntax. Both – built-in NAO system and VoMIS system provides similar functionalities, therefore we provide a short comparison of them.

Built-in system uses ALDialog module for dialogue management. It enables very natural conversations, but it lacks a stronger flow management, which cannot be written directly into the rules and need to be written into the source code. Moreover, writing dialogue interactions for NAO cannot be considered as simple and effective. On the other side, VoMIS system uses VoiceXML-based DM. It enables less natural interaction (mostly system-directed), but it provides a framework for description dialogue flow and content in more details. Moreover, it enables to use more complex speech grammars written according W3C SRGS 1.0 recommendation.

Even though built-in ASR and TTS system supports 19 languages, the Slovak language is not supported. Moreover, some functionalities of ALDialog module and speech+gestures generation are supported only for few languages. VoMIS system enables to use external ASR and TTS modules to support other languages.

VI. Conclusions

The proposed work describes VoMIS multimodal interactive system for NAO robot, where interaction is controlled by VoiceXML scripts. VoMIS system brings an uncomplicated way how to control and maintain human-robot interactions. The undisputed advantage of VoiceXML is its simplicity and readability of dialogue scripts together with a great approach to more structured task-oriented, domain-independent interactions. Moreover, VoiceXML is an industry standard for Spoken dialogue systems, which is widely accepted and well known in the community. We extended the unimodal character of VoiceXML towards multimodal interactions, by overloading the <prompt> element content, which now can bear also instructions for gesture and movements production. The result of our work is the system that enables easily to combine spoken interaction with movements and gestures of the robot.

Moreover it makes possible easily to use external source of speech, external automatic speech recognition and text-to-speech synthesis engines. VoMIS system brings an alternative to the built-in NAO system, which can be more appropriate for dialogue interactions that are more structured and complicated.

Acknowledgment

The research presented in this paper was supported by the Slovak Research and Development Agency under the contract No. APVV-15-0517, APVV-15-0731 and APVV SK-TW-2017-0005.

References

[1] J. Collinászy, M. Bundzel and I. Zolotová, "Implementation of Intelligent Software using IBM Watson And Bluemix," Acta Electrotechnica et Informatica, 2017, 17(1), pp. 58–63. ISSN 1335-8243. DOI:10.15546/aeei-2017-0008

[2] M. Sulír, and J. Juhár, "Hidden Markov Model Based Speech Synthesis System in Slovak Language with Speaker Interpolation," Acta Electrotechnica et Informatica. 2015, 15(4), pp. 8–12. ISSN 1335-8243. DOI: 10.15546/aeei-2015-0029

[3] S. Al Moubayed et al., „Generating RobotAgent Backchannels During a Storytelling Experiment", in: Proceeding ICRA'09 - Proceedings of the 2009 IEEE international conference on Robotics and Automation, 2009, Kobe, Japan., IEEE Press Piscataway, 2009, pp. 3749-3754.

[4] S. Al Moubayed et al., „Multimodal feedback from robots and agents in a storytelling experiment", in: Enterface'08. – Orsay, LIMSI-CNR, 2009, pp. 43-55.

[5] S. Ondáš et al., "Multimodal dialogue system with NAO and VoiceXML dialogue manager", in Proceedings of CogInfoCom 2017, Debrecen, Hungary, 2017, pp.439-444.

[6] S. Ondáš, J. Juhár, „Dialog manager based on the VoiceXML interpreter", In: Proceedings of the DSP-MCOM 2005, 6. International conference, Košice, Technical university of Košice, 2005, pp. 80-83.

[7] Aldebaran Dokumentation [online]., http://doc.aldebaran.com/2-1/naoqi/audio/dialog/aldialog.html

[8] J. Juhár, A. Čižmár, M. Rusko, M. Trnka, G. Rozinaj, and J. Jarina, "Voice operated information system in Slovak,"

Computing and Informatics, 2012, 26(6), pp. 577-603. ISSN: 1335-9150.

[9] S. Ondáš, J. Juhár, M. Papco, M. Trnka and V. Király, "The integration of the Hungarian language in to the Slovak Spoken dialogue system," In Proceedings of the 9th WSEAS international conference on signal, speech and image processing, and 9th WSEAS international conference on Multimedia, internet & video technologies, 2009, Budapest, WSEAS, pp. 102-105