Voice Portals Solution for Medical Consultations Appointment

André Sabino

Department of Information Technology Engineering

Av. Prof. Dr. Cavaco Silva - 2780-990 Porto Salvo, Portugal

andre.sabino@tagus.ist.utl.pt

Abstract: *The Interactive Voice Response systems (IVR) associated potential is big [1] [2] [3], once it is possible to optimise processes, combining existing technology, either in recognition or in synthesis. It acknowledge of a specific area of action, in order to create solutions to ensure the best possible action from the technology.*

Link Consulting is a company that has developed several solutions in the area of voice systems, not only in a different context of what is involved, also with different technologies. It is a reality that once working with people with experience in this area naturally enriches the result of the work.

This document describes the work done, develop a system of voice response, enabling patients to a particular health institution and on this specific case of Social Services Medical Assistance (SAMS), make appointments, using the phone (or mobile) or any application of Voice Over Internet Protocol (VoIP), through voice, or using the keyboard of these devices or applications.

As an innovation associated with this work, besides the technology and what it can withdraw from, compared to others in the market, it describes a rules system, easily configured, these rules can be applied over the working flow regarding this application.

Based on the importance associated with each of the rules, it contributes to the final decision, associated with each of the different phases that already exists in the flow, allowing the optimisation of the functioning base, not just taking advantage of information retained in historical, but also the configuration system so as to take even better advantage of the system in the future, either regarding the time or the success of the work time factor, or factor in the success rate in recognition.

Keywords: *Voice Portal [4], Automatic Speech Recognition, Speech Synthesis, Dialogue Definition Techniques [5], Voice User Interface [6] [7], VoiceXML [8], SALT [9], Windows Workflow Foundation, Dual Tone Multi-Frequency, Voice over Internet Protocol.*

1. Introduction

The term Voice User Interface (VUI) is used to classify the tool that enables in an easy way, to create a natural interaction between humans and machines. In association with the interaction, may be actions and or events that allow the start, the end or the monitorization of various processes, based in the course of that interaction, but in an automatic way.

The technology support for this kind of system exists some time ago, however, can be considered a difficult challenge. The positive acceptance of the VUI depends on the definition and good implementation of the system.

The feeling associated with typical expression of "talking to a machine", is not definitely something that pleases a big number of users. It may occur, prior to that feeling, a bad experience of contacts with such systems, because of failures of implementation, which may not lead to recognition, false recognitions or situations of confusion for users, because of that the dialogue does not properly guided.

With all this, it is possible to reach the conclusion that this type of system has no significant success, because people have some fear of this kind of systems. Also because not always the situations, which intend to implement a system with a voice interface, be duly investigated in order to avoid situations of doubt or uncertainty, either for the system or to the user. This culminates in a bad final event, for an experience that supposed to help and benefit, both sides involved.

Even with high quality in the platforms assignments and on the support of such systems, it is necessary to have the feeling that it is crucial, because it seeks to achieve the success, bringing together different types of abilities, not just for the technology area, but also for a language area and even for the psychological area.
It is important to understand the context in which to apply this system, whether in an environment where users will use a very specific terminology in a way and know how the system works.
In another way, users are dealing with something unknown. The first contact thereby requires the help and the guideless' to conduct to a particular task.

1.1 Problem

On the scope of this work, existing technologies are studied, once allied to the knowledge of the context of use acquired by professionals in a particular area of action, this culminate in a draft solution to the problem that is the implementation of an IVR system that allow users / patients of SAMS, which are more or less experienced, to make appointments for medical consultations, by phone (or mobile) or through the use of any application of VoIP, making use of their voice or of the keys in existing facilities / applications.
As optimization to the system that solves the basic problem, one intends to have any mechanism which aims not only to optimize the normal operation of the system, in terms of time and success rate associated with the recognition, based on past information associated with each user, but also the opportunity to realize parameterizations that the same mechanism, to improve the performance of the system as a whole.
It is important to allow parameterizations, to the rules mechanism, or to some parameters related to the overall system. However, this parameterization could be made by someone without knowledge of technology and should be used as a user-friendly, which allows the configuration of the above-mentioned parameters.
This work aims to extend the system that Link Consulting implemented in this institution, with the goal to provide an alternative to the system of appointments by email and Web browser.

1.2 Adopted Technology

As we have several solutions in the market that allows meeting the requirements imposed on the development and subsequent management of such systems, it is necessary to set some parameters to make a careful selection of the platform to use. Within this respect, some parameters were set to make a pre selection:

- Easy integration of several components required for the functioning of the system (ASR, TTS,…);
- Portuguese language support, with the best possible proportion to the level of recognition;
- Support and cooperation during and after the development;
- Requirements for software or hardware needed to run the system;
- Be an user independent system and allow improvements to the level of recognition;
- Robustness against the noise;

Among the solutions pre-selected offered by Nuance, Philips, Loquendo and Microsoft, the solution that was considered the most viable for the case of this work, was offered by Microsoft, because it allows easily meet some of the taxes goals to the selection of a platform, listed below:

- Easy integration between all the necessary components, including ASR and TTS;
- Be able to guarantee a high percentage of success at the level of recognition, in this case, with Portuguese language;
- Allow developing a scalable solution, which at the same time can be parameterized, with the lowest number of possible implications in the existing structure of the system.

Beyond these points, the following can also be pointed as more valuable to adopt a solution based on managed code, thus weighing on the choice taken:

- Programming Flexibility;
- Easy activation and management of the flow;

• Ability to exit of the execution of the working flow (interconnection with other systems).

2. Specification

This work consists in the development of an IVR system, which allows the users of the SAMS to have another way to make appointments of their consultations, since until now, they only had the following ways: in counter, through the Web browser and through e-mail.
With the design and implementation of the base system, it is also extending through the incorporation of a rules mechanism, allowing the optimization of the functioning of the base system, as already mentioned, the two main levels: the time necessary to complete the task of make an appointment and the success rate associated with the recognition.

2.1 Cases Studies

In order to realize the available possibilities about the operation of the system, three examples are mentioned.
The first one, represents all situations where nothing is known about the historical information of the patient, or even known, information is not used. It is necessary for the user to give all the necessary data to make a new appointment.

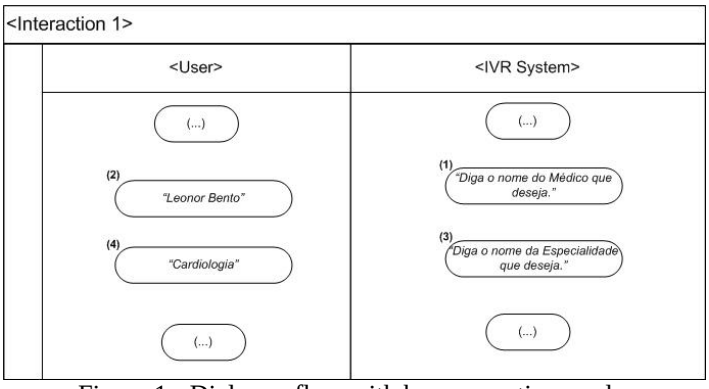

Figure 1 - Dialogue flow with base operation mode.

In the second case, are represented all situations in which, based on the historical information associated with a patient, that it can make a prediction of what he possible says in that interaction, but the percentage of certainty associated with that prediction is not too high, on bringing the formulation of a single question. The user need only to confirm what it is suggested, in each of the different states.

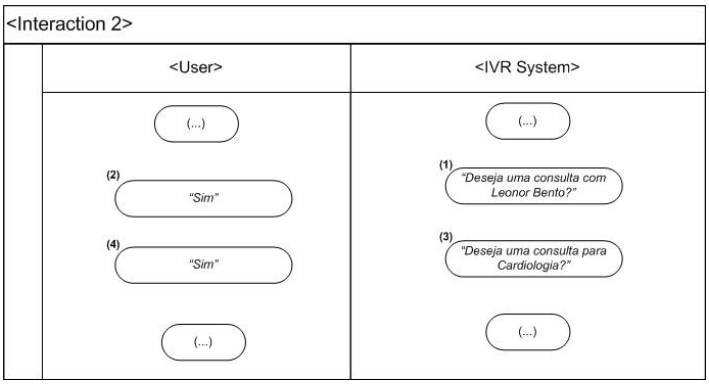

Figure 2 - Dialogue flow with optimized operation mode (level 1).

In the third case, are represented all situations in where historical information is used. This can be a relatively high percentage of certainty, given what the patient possible says on that interaction, thus, make a single question. The patient just needs to confirm a single question.

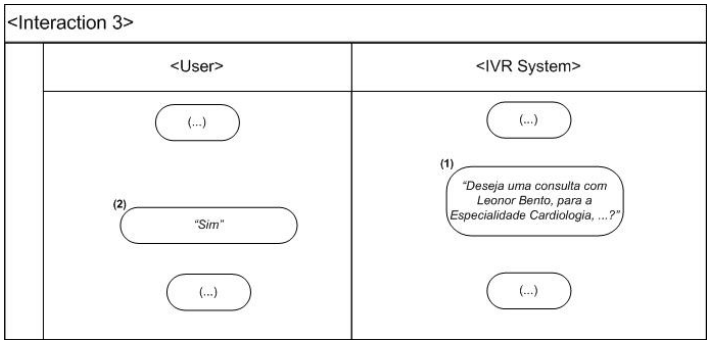

Figure 3 - Dialogue flow with optimized operation mode (level 2).

2.2 Some Considerations

Talking about dialogues definition, there are some relevant points that should be taken into consideration, among which some are listed bellow:

- Having notion of the capacity of human memory;
- Design and concept of what is the end user, knowing the universe of users;
- Let the user lead the conversation or have that feeling, even if it is not entirely true;
- Keep going on the work flow and avoiding to return;
- Designing the dialogue based on the largest group with common interests, not harming the majority as a benefit of some;
- Allow the system to do some work:
    - In terms of the working flow, not getting the information only through questions, but also in other ways;
    - In terms of the system recognition, using small grammars, either static or dynamic, list of potential recognition, clear definition of *prompt's*.

2.3 Rules Mechanism

The rules mechanism, whose implementation was proposed, incorporates something that allows the optimization of the flow basis application, based on rules that can be applied during the various stages of that same flow. This mechanism should allow the operating system in its basic version, or a version optimized, and thus configurable to this and other levels.
The introduction of this mechanism, should work as the incorporation of a new component to the basic version of the system, and hence little structural change in the application.
As an associated to the mechanism, it is considered the possibility to configure those same rules, thus making it possible to adapt the general system in something adaptive, once is now possible to directly influence the operation of the system through this same rules mechanism.

As quoted before, the main factors, which aim to optimize, the incorporation of a rules mechanism are:

- The time required to accomplish tasks

One of the main points to optimize is precisely the time needed to make a new appointment for a consultation.
Among the factors that lead to attempt to meet this objective, are monetary issues, or even issues related to the availability not only of users, but the system itself.

- The rate of success associated with the recognition

Another key factor, which aims to achieve, is a good percentage of success, in association with the recognition.
In order to meet this requirement, and try to combine the best practices within the definition of dialogues with the available opportunities, on the technology used.
The use of grammars, which contain the essential information, is one of the paths to follow, which can be useful. Those grammars have not always the same size or the same set of words and can therefore be

designated for Dynamic Grammars because they can be modified over the flow of execution of the application.

Let's suppose that a user connects to the appointment system, and that the system might discover that with a certain percentage of 100%, the user will try to make the appointment of a consultation with "Leonor Bento" and, as attached as in the example make a direct question, "Do you want a consultation with Leonor Bento? ", to which the user answered "Yes". In a situation like this, it does not make sense, the next question involved in the flow of execution, "Which Specialty do you want?", has as possible answers, all specialties that exist in SAMS, but only those with whom a relationship can be established, with the doctor, previously confirmed.

The rules mechanism is what will make the base system possible, to behave in one way or another, depending on the information that is generated by the same mechanism.
The rules mechanism "suggests" what is the best way forward from the current state in the flow of execution, not only based on a specific user/patient, but also based on the historical analysis of information, associated with the same user/patient.

Assuming that once set the system to make use of the rules mechanism, in addition to operating the system, the first information that is intended to confirm the user is what <Doctor> to which he intends to go.
Associated to the patient concerned, there is a history of information, that indicates all consultations with their doctors and specialties, locations and times, associated with these consultations.
Assign up then by state, each point along the flow of execution of the application, where it is seeking "critical information ", i.e. <Doctor>, <Specialty>, <Locality> or <Time>.
Each rule is linked to a state, thus indicating that the same state, think about what is the most probable hypothesis, which aims to meet the needs of the patient.
Since there may be several rules associated with the same state, it is necessary to assign weights to them, or what percentage / weight given to his suggestion, when it is grouped with all other suggestions, for a decision end for a state.

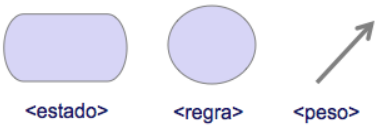

Figure 4 - Basic elements to conceptually illustrate the rules mechanism operation.

<State/Estado>: represents a specific state of the flow of execution of the application, in which there is a problem to solve, i.e. obtain "critical information ".
<Rule/Regra>: process that helps to solve a particular problem, coupled with a <State>.
<Weight/Peso>: input (%) of a certain <Rule> for the resolution of a problem associated with a <State>.

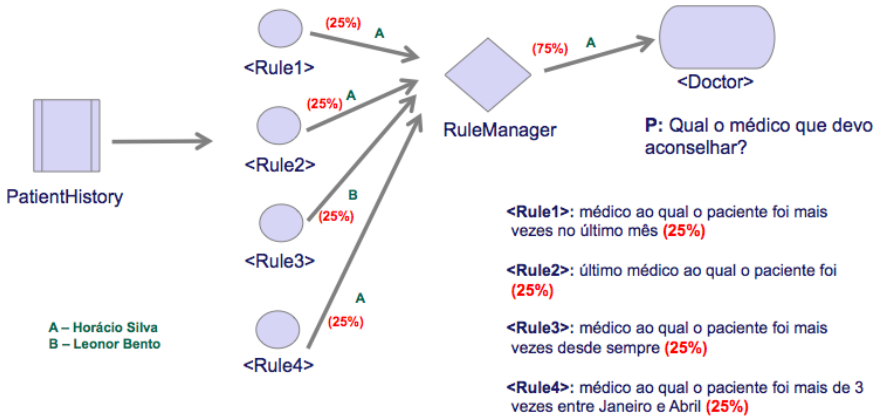

Figure 5 - Operation of the rules mechanism in the state where is needed information about the Doctor.

In the previous example, to the state <Doctor>, four rules were defined, all with equal levels among themselves once one can think that the suggestion given by each of them, for the final decision, has equal importance, whether or not being the same.
One can verify that with the historical analysis performed by each of the rules in order to obtain the information on each of them have, three rules has the same suggestion, including the Doctor A, making a unanimity among 75% of shares.
Whereas 75% an acceptable weight, is then taken a final decision, and finally suggested the Doctor A, as the best option, and that most likely will in fact what the patient demand.

With this rules system, there is a real possibility of having combination of different modes of operation, to collect as soon as possible, all necessary information to create a new record. Assuming that the user only agrees with part of the information it is suggested by the system, for example, only want the doctor and specialty suggested. In case of the user does not accept what suggested, there is a passage from optimized working flow, in this case the optimized flow (level 1) for the basis flow, without losing the information that has already been confirmed by the user, continuing from there.

3. Development

3.1 Technology

As already stated above, the technology used in this work, it is all provided by Microsoft, offering some facilities to define the dialogues, after the installation of additional modules and plug the basic platform, which allows us to optimize the process of development of such applications.

As active parties and essential components of this work, there are two fundamental entities that can be used by developers:

States / Activities

It consists on a Speech Application, which states along the application are defined using the features incorporated in Visual Studio 2005, both in terms of simple Speech Applications, or through integration with the extensions above, allowing this way to define the entire execution flow, representing each state of that flow, as an activity.
The use of workflows, gives the possibility to do the wrapping of components of voice response, in activities which do not only provide an abstraction on the code that is executed, but also because it allow reuse those components.
The workflow can be seen as a hierarchical structure of activities. Such activities may be simple (primitive), or become containers for more activities.
The layer of Dialog Activities of Speech Server API offers a series of classes of primitive objects, which include the following:

- *QuestionAnswerActivity* - implementation of question-answer;
- *StatementActivity* - single command execution;
- *RecordAudioActivity* - enforcement of a prompt, recording with the response of the user.

In addition to these activities, others are available in order to allow a certain level of abstraction in the face of the primitive voice response, and too make possible the reuse of components over the flow of execution.

Grammars

Depending on the type associated with each state, or there may be no need for the creation of grammars, whether for use in speech recognition or in DTMF recognition.
There is also the possibility of creating grammars using the Grammar Builder, the Grammar Creator or combinations of the two mechanisms.
The creation of dynamic grammars, defined as the application is executed, as appropriate for a given state, is also a possibility.

3.2 Dialogue Definition

The following extract of code, shows how well prepared the welcome message is, when the user tries to use the system.
As a matter of monitoring and evaluation to the functioning of the system, is used a log mechanism which may be found in the management application. It is possible to understand what the general flow of execution in each interaction with users of the system, which rules that were applied and the reason for their application, among other information related to the overall functioning of the system.

```
private void Welcome_TurnStarting(object sender, TurnStartingEventArgs e){
       this.Welcome.MainPrompt.ClearContent();
       this.Welcome.MainPrompt.AppendText("Bem-vindo ao sistema de
marcação de consultas do SAMES.");
       _context.logger.WriteMessage("[" +
       this.ApplicationHost.TelephonySession.Id + "] " + "[WELCOME]");
}
```

Figure 6 - Extract of code representing the Welcome state.

The following extract of code is part of the process of validating the identity of a user who tries to access the system. Once the required number of beneficiary and the personal code, it is trying to obtain the data source, a registration number of which is equal to the beneficiary given by the user. If there is user and given code is correct, the validation is done.

```
NewPatient = _context.GetPatient(PatientId);

if ((NewPatient != null) && (PatientCode == NewPatient.CodePatient))
{
   this.PatientExists = true;
}
else {
   this.PatientExists = false;
}
_context.logger.WriteMessage("[" +
this.ApplicationHost.TelephonySession.Id + "] " + "[USER VALIDATION]");
_context.logger.WriteMessage("[" +
this.ApplicationHost.TelephonySession.Id + "] " + "[Patient Id: " +
PatientId.ToString() + " Password: " + PatientCode.ToString() + "]");
```

Figure 7 - Extract of code used to validate the information given by the user.

One time the user is logged to the system, it is time to be given the options through which to start the process of make a new appointment. Since you can get by Doctor or Specialty, it is necessary to understand what it stated, in order to define what should be questioned then. The part of the general workflow, which is associated with this stage of interaction, is shown in the following image.
Under these states, is the application code to understand if the user said a Doctor name or a Specialty name.

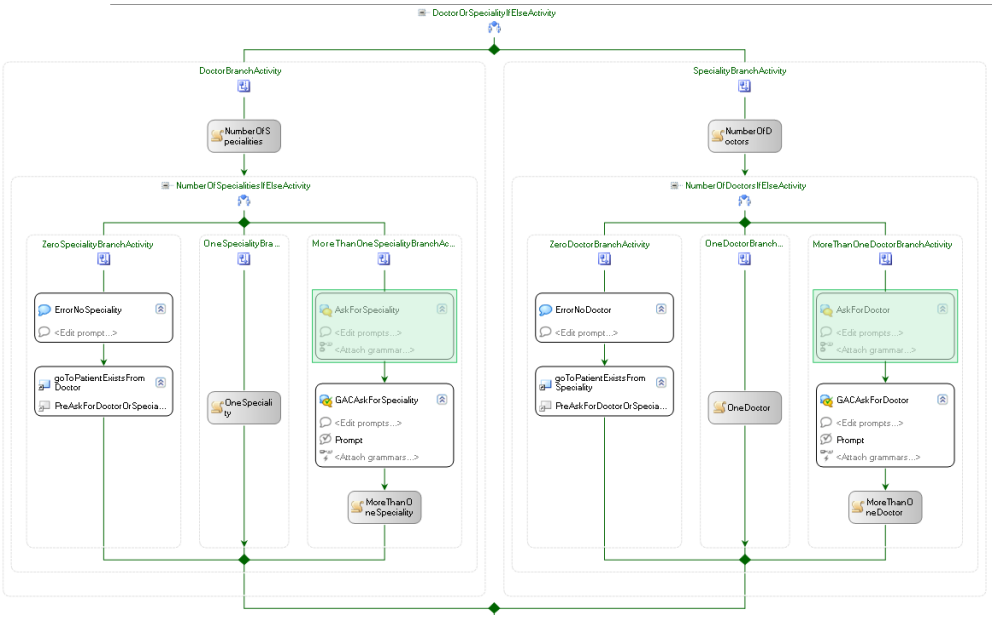

Figure 8 - Request for Doctor or Specialty.

The application has five universal commands, which give access to the help system, which allows repeat a prompt, which returns to the beginning of interaction with list of menu options, which returns to the state where choose a doctor or a specialty and the command that allows end the interaction with the system.
Of the five commands above, the command that gives access to the help system and command that allows repeat something said earlier by the system, in the state where the interaction is, are dependent on the context, either because the help given to the user by the system, either to repetition of something that was said before, and that for some reason the user intends to hear again, depend on the state where the interaction is. It is therefore necessary to know the constant evolution of each interaction.
In this sense, it is necessary to define the prompts *RepeatPrompt* and *HelpPrompt* so that whenever one of the universal commands is invoked, the prompts can be used.

The commands that allow the go back on the interaction and the command that allow go to the end of the interaction, are not dependent on the context, and is thus easier to implement, since regardless of the state where the user is along the interaction, action in will always hold a *go to*, for a previously defined state, be it for the listing of options, to select a doctor or specialty or to prompt the farewell. Because of the facilities offered by extensions to WWF, we can do so using the entities already defined, with only the necessary changes to some parameters. The image shows an existing activity, configured to pass to the state Goodbye, when you request the termination of the interaction.

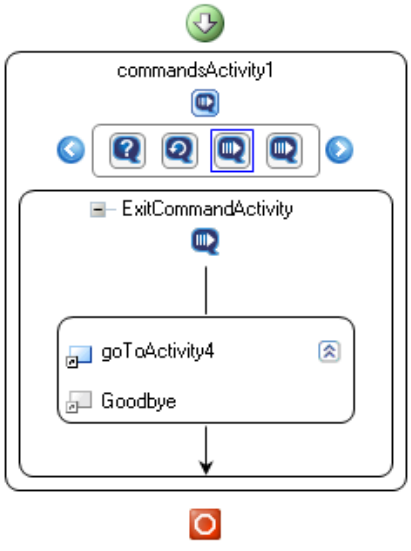

Figure 9 - Definition of the universal command Exit / Goodbye.

4. Results and Evaluation

To assess the system, in situations as close to real as possible, were set various scenarios to test, in several different environments, as it is considered of great interest, plan and evaluate properly this type of systems [10] [11] [12], whose operation can vary, even if the "entries" not be significantly different.
Experienced users tested the system, in order to create some benchmarks, which could then provide, based on each of the tasks, the time that normal users, would result in the achievement of these tasks.
Then an analysis will be carried out with results and drawn some conclusions about the same.
The completion of these tests for the operation of the system, which intends to evaluate the optimum that achieves the level of two key aspects, the time needed to perform the tasks, in this case, make a new medical consultation appointment in various environments, but also to optimize the level of the success associated with the recognition.

In order to evaluate the system, in situations as close to reality as possible, were set to test various environments, since such a system should be used in any environment, regardless of existing noise.
Since there are several operation modes, as noted above, the situations of tests offered to users, were to combine the environments and operation modes, drawing conclusions on each of the interactions made.
After the tests, it is possible to present some results for optimizations achieved, either at the time necessary to perform the tasks, even making the interaction vary between different modes of operation and different environments for testing. The following table shows in simplified form, the possible gains in terms of time.

Table 1 - Gains in various combinations of Environment and Operation Mode.

	Mode 1	**Mode 2 (1 → 2)**	**Mode 3 (2 → 3 / 1 → 3)**
Environment 1	142s (100%)	97s (~ -31.7%)	74s (~ -23.7% / ~ -47.9%)
Environment 2	160s (100%)	104s (-35%)	81s (~ -22.1% / ~ -49.4%)
Environment 3	233s (100%)	111s (~ -52.4%)	85s (~ -23.4%/ ~ -63.5%)

The following graph represents the average proportions, attached to successfully achieved recognition, recognition with confirmation by the user and not recognition, being equally vary the interactions among different operation modes and different environments.

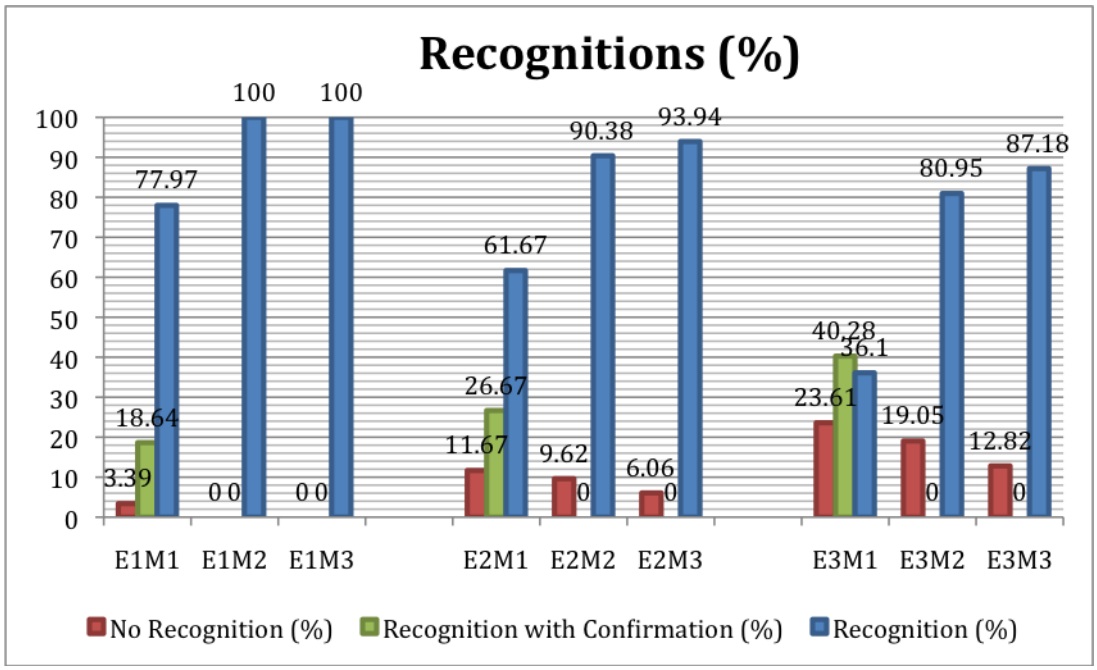

Figure 10 – Recognition, Recognition with Confirmation and No Recognition.

5. Conclusion

5.1 Work Done

After this work, is a great satisfaction to see that the requirements set for it, have been met and that thus been achieved design and implement an IVR system, which allows the marking of medical consultations, in this case to the SAMS, using the technology made available to the Link Consulting. With no less satisfaction, finding that although taxes have not been very specific requirements as to the rules mechanism, which aimed to optimize the system, in its basic version, has been developing a module, which may or may not be used with the basic version of the system, easily parameterization, which can in some cases, optimizations to about 50% reduction of time to perform the task, as well as associated rates of 100% in recognition with success.

To assess the solution presented, tried to take the same tests, which correspond as closely as possible to existing reality, when we talk of a consultation marking system, through an IVR. It was of tremendous value, there are people willing to spend more than thirty minutes, to make possible test the system in various environments and through various optimization levels. A situation of such tests, allows not only understand the correct or incorrect operation of the system, but also understand the extent to which technology is used properly mature, compared to the other referred before.
In addition to the lessons that can be taken personally, there is an increase in benefits to test such a system, with several people, and for some time with each person, corresponding to the notes given by each one of them, while testing the system, which can lead to the realization of adjustments to the system itself.

After this work, there is a tested version, which complies with the requirements set out initially, to allow perform an assessment to the system, when it was finished. Thus, this work can be seen as a starting point for further work in this area, with the technology used, because few projects are known, made public, using these technologies from Microsoft.

5.2 Future Work

Integration Foreign Rules System - BizTalk Business Rule Engine (BizTalk BRE)

The integration of the solution, with external rules systems as the BizTalk BRE, will give the solution a higher degree of extensibility, since it becomes the BRE responsible for the decision to take, based on the results of the implementation of various rules defined in the system, making it even possible to manage own rules, through the interface provided by BRE.
Besides the fact that it is possible to use BizTalk BRE to manage the rules, it will be possible too, make available the same rules for other uses, in the same context.

Embed updates to the Speech Server 2007

The incorporation of updates, either at the recognition system, either at the synthesizer, the Microsoft Speech Server 2007, will significantly improve the outcome of any application using this technology.
The rate of recognition will move to more acceptable levels, which together with good practice for the dialogues definition could give a final higher level of satisfaction, compared to the present. The improvements in the recognition system, due to the acquisition of records of calls, will enrich the vocabulary used by recognizer and thus improve the rates of recognition successfully.
The improvements of the synthesizer are the incorporation of a more natural voice, compared to currently used by the platform, which will also contribute in some way, to increase the level of satisfaction of users.

6. References

1. **Gunar Fiedler, Peggy Schmidt.** *Developing Interactive Voice Response Interfaces for Large Information Systems.* [Paper] Institute for Computer Science and Applied Mathematics, Germany.
2. **Rudžionis, A., K. Ratkevičius, R. Maskeliūnas, V. Rudžionis.** *Review of Voice Dialogues in Telecommunications.* [Paper] Speech Research Laboratory, Kaunas University of Technology, Dept.of Informatics, Kaunas Humanities Faculty of Vilnius University. Kaunas, Lithuania.
3. **Schoelle, Art.** *Voice Self-Service Aligns with Web Architectures to Reduce TCO, Increase Flexibility and Ensure Content Consistency*. [Paper] Customer-Centric Strategies. Yankee Group. 2006.
4. **Steinbiss, V..** *Speech Dialogue Systems In The Telecom Network.* [Paper] Philips Speech Processing Kackertstr. 10, 52072 Aachen. Germany.
5. **Fiedler (1), Gunar, Bernhard Thalheim (1), Peggy Schmidt (1), Thomas Schwanzara-Bennoit (2).** *State-, Hhml-, And Object-Based Dialog Design For Voice-Web Applications.* [Paper] (1) Computer Science Institute, University Kiel, Germany. (2) Computer Science Department, Databases and Information Systems Group, Brandenburg University of Technology at Cottbus, Germany.
6. **Schnelle, Dirk, Fernando Lyardet.** *Voice User Interface Design Patterns.* [Paper] Teleccoperation Group, Darmstadt University of Technology, Darmstadt, Germany.
7. **Mecanovic, Daniel, Hao Shi.** *Voice User Interface Design for a Telephone Application Using VoiceXML.* [Paper] School of Computer Science and Mathematics, Victoria University. Australia.
8. **VoiceXML Forum.** *Voice eXtensible Markup Language – VoiceXML Specification.* [Paper] VoiceXML Forum.
9. **SALT Forum.** *Speech Application Language Tags (SALT) Technical White Paper.* [Paper] SALT Forum.
10. **Lane, Ian R., Shinichi Ueno, Tatsuya Kawahara.** *Cooperative Dialogue Planning with User and Situation Models via Example-based Training.* [Paper] School of Informatics, Kyoto University Yoshida-Hommachi, Sakyo-ku, Kyoto 606-8501, Japan.
11. **Walker, Marilyn A., Diane J. Litman, Candace A. Kamm, Alicia Abella.** *Evaluating Interactive Dialogue Systems: Extending Component Evaluation to Integrated System Evaluation.* [Paper] AT&T Labs Research, Florham Park. USA.
12. **Facco (1), A., D. Falavigna (1), R. Gretter (1), M.Viganò (2).** *On the Development of Telephone Applications: Some Practical Issues and Evaluation.* [Paper] (1) ITC-Irst, Istituto per la Ricerca Scientifica e Tecnologica, Trento. (2) Reitek Spa, Milano. Italy.