# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

**4,800**
Open access books available

**122,000**
International authors and editors

**135M**
Downloads

Our authors are among the

**154**
Countries delivered to

**TOP 1%**
most cited scientists

**12.2%**
Contributors from top 500 universities

CLARIVATE ANALYTICS BOOK CITATION INDEX INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

## Interested in publishing with us?
## Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

# 17

# Development and Evaluation of the Spoken Dialogue System Based on the W3C Recommendations

Stanislav Ondáš and Jozef Juhár
*Technical University of Kosice, Faculty of Electrical Engineering and Informatics*
*Slovakia*

## 1. Introduction

Due to progress in technology of speech recognition and understanding, the Spoken dialogue systems (SDS) have started to emerge as a practical alternative for a conversational computer interface. They are more effective than Interactive Voice Response (IVR) systems since they allow a more free and natural interaction. The Spoken dialogue systems are designed for providing automatic dialogue-based voice services accessible through telephone. Such systems consist of a number of components that need to work together for the system to function successfully (McTear, 2005). The basic architecture of the SDS consists of (more or less indispensable) modules – dialogue manager, language understanding, speech recognition, access device interface, language generation and text to speech synthesis. Easiness of implementation and rapid development of voice services led to the standardization effort. The World Wide Web Consortium plays an important role in this area.

The book chapter proposed will be focused on design, development and evaluation of the Spoken dialogue systems based on the W3C Recommendations. The World Wide Web Consortium is an international community that develops standards to ensure the long-term growth of the Web (W3C, 2010). One of their workgroups, Voice Browser Working Group, deals with preparing standards for voice-enabled technologies. The main idea is to build "Voice browser" enabling access to the information by voice, similarly as in the case of web browser. Comparison of definitions of the Spoken dialogue system and Voice browser lead to the conclusion, that both systems are very similar or de facto identical. A group of XML-based languages (SIF - Speech Interface Framework) was defined by Voice Browser Working Group to enable speech communication between user and computer. The W3C SIF recommendations became the industry standards in voice-enabled technology domain during the last decade. Languages in SIF also define the interfaces between fundamental subsystems of Spoken dialogue system and thus determine the basic structure of such system. The main languages in the framework are VoiceXML, SRGS and SSML that enable composing dialogues, speech grammars and instructions for text-to-speech systems. The CCXML serves for handling I/O (telephony) devices. The SISR specification defines the semantic tags for speech grammars to enable extracting of the meaning of user's input. The meaning can be represented in the EMMA language, which was prepared by the W3C

Multimodal Interaction Working Group (MIWG, 2010). The PLS defines tags for composing pronunciation lexicons for ASR as well as TTS systems.
Evaluation of the Spoken dialogue systems and their services is also very important in the system's life-cycle. During the test phase it may bring the information about the performance of the system, in the phase of pilot running enable to observe the impact of changes, which were done and in the phase of public running can provide an estimation of the quality perceived by the users.
At the beginning of the article we will provide description of these languages and some other important technologies and then we will outline the architecture of SDS, which is based on W3C recommendations. Then, in section three, the research and development of the Slovak spoken dialogue system (SDS) will be described, which adopts several W3C languages. The architecture and components of the system will be introduced. The last part of the article will be focused on objective as well as subjective evaluation methods. Both methods bring different information about the system and services being provided. The objective method based on collecting of interaction parameters will be presented. For that purpose the evaluation server was integrated into the Slovak SDS. It is described in subsection 3.2.6. Also a subjective evaluation based on filling in the questionnaire was carried out and will be described in the section four.

## 2. Description of the W3C Speech Interface Framework

The World Wide Web Consortium (the W3C) is an international community developing the standards ensuring the long-term growth of the Web (W3C, 2010). The Consortium consists of working groups associated with the research area. One of them is the Voice Browser Working Group (VBWG) focused on development of standards for Voice Browsers. Voice browsers allow people to access the Web using speech synthesis, pre-recorded audio, and speech recognition through their phone device. The Voice Browser Working Group was first established on March 26, 1999 (VBA, 2010), to develop specifications for these devices. The W3C Speech Interface Framework (SIF) is a suite of markup specifications aimed at realizing this goal. Languages in that group fulfil the idea of portability and rapid development of voice services. The framework actually consists of VoiceXML, SRGS, SSML, PLS, SISR, CCXML and SCXML specifications. Following subsections provide their short description.

### 2.1 Voice eXtensible Markup Language

The VoiceXML (Voice eXtensible Markup Language) is a markup language designed for composing the voice applications. In 1999 four companies AT&T, IBM, Lucent and Motorola established the VoiceXML forum (VXMLForum, 2010) for designing a language, which would increase the development of voice applications. The first version of the language was introduced in august 1999. The first official version of the VoiceXML language (VoiceXML 1.0), prepared by VoiceXML forum, was presented in March 2000. After that the W3C adopted the responsibility for VoiceXML language, and it had started working on the next version of VoiceXML.
Whereas the VoiceXML 1.0 language specification implied, besides tags (markups) for dialogue description, also tags for call management, speech grammars and speech synthesis, the second version of the language (VoiceXML 2.0) was focused only on dialog description. Markups for call management, speech grammars and speech synthesis control were adopted as the background for CCXML, SRGS and SSML languages. VoiceXML 2.0 was released as

the W3C recommendation in March 2004. This recommendation became the industry standard in area of voice services.
In June 2007 the VoiceXML 2.1 was introduced, which attaches a tiny set of additional features to the second version of the language. Then working on the new specification (VoiceXML 3.0) has started, with the concept of three layers - dialog, flow and management. Work on the third version of the language is still in progress.

### 2.2 Speech recognition grammar specification

As mentioned above, a tiny set of markups used in VoiceXML 1.0 language has created a base of Speech Recognition Grammar Specification (SRGS). SRGS specification brings a language, which enables arranging context-free grammar for speech or DTMF input. Grammar can be specified in either XML or an equivalent augmented BNF (ABNF) syntax. Work on this language has been started in 1999 and it became the recommendation in March 2004 (SRGS 1.0).
The main advantage of the SRGS is well readable form both for designers and computers. It enables composing possible language structures, that are expected from user in actual state of interaction (dialog). Creation of such structures helps the speech recognition system to be more accurate and faster.
SRGS specification can describe (handle) also speech input in a form of utterances in natural language, but it does not support stochastic language models (N-grams) directly. The N-gram specification serves for that purpose, but it has never been published as the W3C recommendation and its preparation did not continue.
The power of SRGS specification is in cooperation with the next W3C specification – Semantic Interpretation for Speech Recognition (SISR).

### 2.3 Semantic interpretation for speech recognition

The semantic interpretation specification describes annotations to grammar rules for extracting the semantic results from recognition. This provides markups and attributes, which can be included in to context-free grammar and thus some semantic information can be extracted by interpretation of these markups. De facto, it does not really "understand", but it is the acceptable approach to the interpretation of spoken language. Such approach can be used also with the input utterances in natural language. In this case, the system can be viewed like keyword-spotting system. It enables capturing keywords in a natural language utterance and assigning them some semantic value and creating pairs of keywords and their semantic values. This concept is very powerful in domain-specific voice services, but almost unusable in communication with conversational agents.
Work on this specification had started in April 2003 and in April 2007 it became the W3C recommendation.

### 2.4 Pronunciation Lexicon specification

Pronunciation Lexicons describe phonetic information for use in speech recognition and synthesis. The requirements were first published on March 12, 2001, and updated on October 29, 2004. The pronunciation lexicon is designed to enable developers providing supplemental information on pronunciation for items as are place names, proper names and abbreviations. The W3C Recommendation was published in October 2008. Such lexicon can be used both by automatic speech recognition systems and text-to-speech systems.

### 2.5 Speech Synthesis Markup Language

As in the case of SRGS specification, designing the Speech Synthesis Markup Language (SSML) has started in year 1999. This process led to the first recommendation of the language (SSML 1.0) in September 2004. Work on this language is still not finished. The SSML 1.1 specification provides a tiny set of additional features to make this language more usable. The speech synthesis specification (SSML) defines a markup language for prompting users via a combination of pre-recorded speech, synthetic speech and music. It provides uniform API between voice platforms and Text-to-Speech engines and enables changing voice characteristics, like gender, speed, volume, etc.

### 2.6 Call Control eXtensible Markup Language

The W3C is designing the Call Control eXtensible Markup Language (CCXML) to enable fine-grained control of speech (signal processing) resources and telephony resources in a VoiceXML telephony platform. CCXML is designed to manage resources in a platform on the telecommunication network edge. It can handle actions like call screening, call waiting/answering and call transfer.

Requirements for that language were prepared in April 2001 and now the language has status "recommendation candidate" (April 2010). This specification brings very important unification into the call traffic handling, because of large range of telephony hardware producers. It releases voice services designers from concerning about hardware-specific application interface and it gives them the high-level interface by the CCXML language.

### 2.7 State Chart eXtensible Markup Language

A State Chart XML or the *State Machine Notation for Control Abstraction* is the last part of the Speech Interface Framework. SCXML is a candidate for being the control language within VoiceXML 3.0, the future version of CCXML, and the multimodal authoring language. Its development started in July 2005 and currently the seventh working draft was published. This new specification is connected to the new idea of *data-flow-management* framework. The main idea is separation of these three layers, because of higher transparency.

### 2.8 Extensible MultiModal Annotation markup language

The Extensible MultiModal Annotation markup language (EMMA) is a markup language intended for use by systems that provide semantic interpretations for variety of inputs, including but not necessarily limited to, speech, natural language text, GUI and ink input (MIWG, 2010). It provides a group of tags for describing semantic of such inputs. The language is developed by the *W3C Multimodal Interaction Working Group* (MIWG, 2010), which aims at developing specifications to enable accessing the Web using multimodal interaction. The first working draft for this specification was published in August 2003 and it became the recommendation in February 2009.

### 2.9 The W3C-based architecture of the voice browser

The W3C Speech Interface Framework languages have their main employment in voice browsers as well as in spoken dialogue systems. The languages from SIF determine the key ideas about cooperation between voice browser components; de facto they determine the architecture of such system.

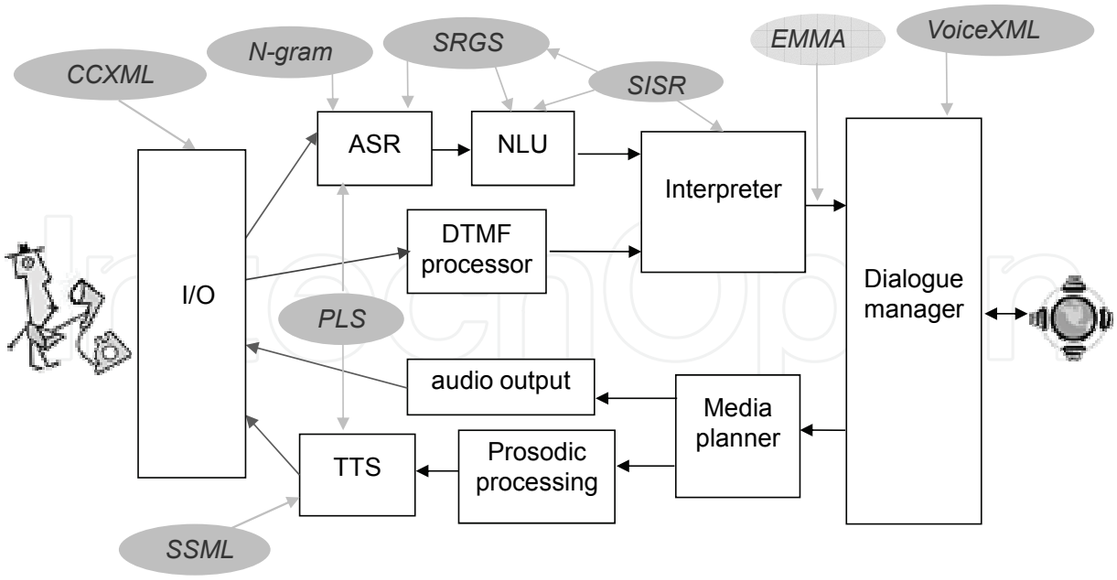



Fig. 1. General structure of the voice browser architecture.

At the end of year 1999 Voice Browser Working Group published working draft of document "Model Architecture for Voice Browser Systems" (MAVBS, 1999). Authors of the document allege that they only wanted to illustrate one of possible solutions of the Voice browser architecture and that the other types of architecture can be adopted for implementation of Speech Interface Framework languages. Interfaces between voice browser's components were not specified directly. The languages within the SIF determine the way of communication between them, what is the main advantage of the Speech Interface Framework. The model architecture of Voice browser from document mentioned above, redrawn into well arranged form in (Delgado, 2005), with stand-alone I/O component is displayed on Fig. 1.

The shadow ellipses represent SIF languages, which should be supported by voice browser components. The main components of the browser are dialogue manager, automatic speech recognition system (ASR), text-to-speech system (TTS) and Input/Output component. The NLU component is responsible for extracting the meaning from user's input. DTMF processor enables processing the DTMF input and both types of input (speech/DTMF) are finally processed in the Interpreter. On the other side, there are Media planner component and audio output block. A block of the prosodic processing is often the part of TTS system.

## 3. The W3C based Slovak spoken dialogue system

The Slovak spoken dialogue system has been developed in period from July 2003 till June 2006 and was supported by the National program for R&D "Building of the information society". The main goal of the project was the research and development of the SDS for information retrieval using voice interaction between humans and computers. The SDS had to enable multi-user interaction in Slovak language through telecommunication networks and to find information distributed in computer data networks such as the Internet. The SDS is also a tool for continuous research in the area of spoken language technologies in Slovakia (Juhár et al., 2006).

The choice of the solution sourced from contemporary free resources, state-of-the-art in the topic and the experiences of the partners involved in the project. Portability and easiness of compiling new services were considered as important factors. The final solution is based on the DARPA Communicator architecture with the central hub process, a software router developed by the Spoken Language Systems group at MIT, subsequently released as an open source package in collaboration with the MITRE Corporation, and now available on SourceForge (Polifroni & Seneff, 2000). The architecture of the system was designed to be compatible with the W3C Speech Interface Framework. The proposed system consists of a Galaxy hub and six modules (servers).
Since 2006 the Slovak SDS is being improved continually at Technical University of Košice with collaboration of Slovak Academy of Science in Bratislava.

### 3.1 System architecture

The architecture of the developed system uses a 'hub-and-spoke' architecture: each module seeks services from and provides services to the other modules by communicating with them through a central software router - the Galaxy hub. Mentioned system (Fig.1) consists of a hub and six system modules: telephony module, automatic speech recognition (ASR) module, text-to-speech (TTS) module, back/end module (Information server), module of dialogue management and the evaluation module. The relationships between the dialogue manager, the Galaxy hub, and the other system modules are represented schematically in Fig. 1.

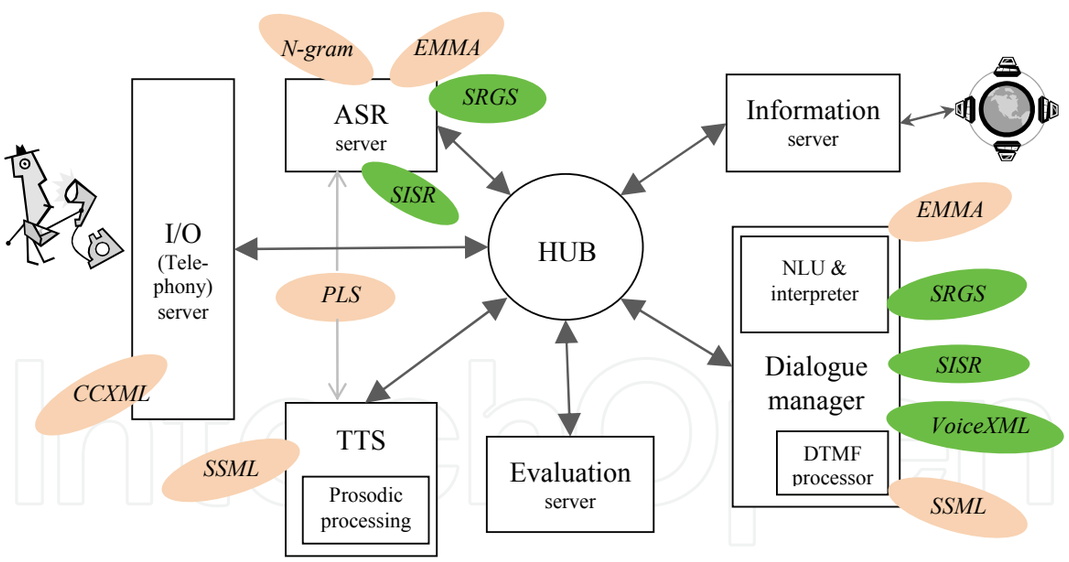



Fig. 2. Architecture of the Galaxy/W3C based Slovak spoken dialogue system

The white rectangles depict the system modules. Green ellipses show the W3C languages, which are supported by Slovak SDS. The orange ellipses show the W3C languages, which can be implemented for full support of the W3C SIF. As we can see on the Fig. 2 some specifications must be supported by more than one system module. For example the SRGS specification must be supported by ASR server as well as Dialogue manager. Dialogue manager needs to generate SRGS grammars and needs to analyze them. The ASR server also needs to analyze SRGS grammar and use it in the recognition process.

As was said above the principle of communication in proposed architecture is the exchange of messages, which are requests on specific functionalities or replies to those requests. The key functionality of messages is to establish the connection, to create and transfer requirements on services of other servers, to transfer replies to those requirements, to inform other servers of the system and so on. In the SDS based on the W3C SIF languages, messages are also the way of transporting code of the SIF languages (part of codes, files, addresses), that hold all information, which is required for communication between servers. From that point of view, messages are only transporters of communication.

### 3.2 System modules

Proposed solution of the SDS system poses common requirements on the module's structure. Each module should have some sub-modules, specifically Galaxy interface, XML Parser and Resource Manager. The Galaxy interface is responsible for communication between module and hub process. It defines dispatch functions (functions, which are provided by the module to the system) and it handles incoming and outgoing messages. Almost all modules need to analyze XML-based documents, because the W3C languages are XML-based. Therefore an XML Parser is the indispensable sub-module. The last one is the Resource manager sub-module, which handles all resources (source files, audio files, models, etc.).

#### 3.2.1 Telephony (I/O) server

A telephony module connects the whole system to telecommunication network. It opens and closes telephone calls and transmits speech data to/from the ASR/TTS modules through the broker channel. The server supports telephone hardware - Dialogic D120/41JCT-LSEuro voice board (Juhár et al., 2006), which creates an interface to PSTN and GSM network (through hardware GSM gateways). Nowadays VoIP (Voice over IP) technology is becoming widely used in telecommunication. The key role is played by the Session Initialization Protocol (SIP), developed by IETF, which is a text-based protocol, similar to HTTP and SMTP, for initiating interactive communication sessions between users (Rosenberg, 2006). The Slovak SDS is connected also to VoIP network by integrating the Open Source library PJSIP in the telephony module (Pleva et al., 2008). The server is ready for integrating of CCXML language, which should be responsible for the interaction between SDS and the telephony module and for the management within the module.

#### 3.2.2 Automatic speech recognition server

The automatic speech recognition (ASR) server performs the conversion of incoming speech to the corresponding text. The ATK based speech recognition engine was adopted for our SDS. The ATK is an Application Toolkit for HTK, which is freely available for non-commercial research (Young, 2004). The Audio input module of the basic recognition system was substituted by the new one, with the interface attached to the Galaxy broker channel. Also there were functions added for creating a group of language sources (grammars and dictionaries), for converting the grammar format from SRGS to HTK-compatible format and backwards, for indicating of VoiceXML *noinput* and *nomatch* events etc. The ASR server in the Slovak SDS supports SRGS and SISR specifications.

Context dependent HMM acoustic models trained on SpeechDat-Sk (Pollak et al., 2000) and MobilDat-Sk (Rusko et al., 2006a) speech databases were used for recognition of the Slovak

language. Context dependent (triphone) acoustic models were trained in a training procedure compatible with "refrec" (Lihan et al., 2005).

### 3.2.3 Text-to-speech server

The text-to-speech (TTS) synthesis server converts outgoing information having the text form into the acoustic form. A concatenative text-to-speech synthesis engine for Slovak language was developed by Slovak Academy of Science. Diphones were selected as a good candidate at this type of synthesis for Slovak language. Two unit selection algorithms that create final sequences of diphones were prepared. Both are based on minimizing the number of artificial concatenations in a synthetic signal. The first algorithm propagates from the longest units to the shortest ones, first querying the corpus for the whole phrase, then for its sub-phrases and finally for the diphones. The second algorithm starts from the shortest units. It just queries the corpus for all of the diphones that are then put into a lattice of candidates. Finally Viterbi search is used to find the path through the lattice with minimum number of concatenations (Rusko et al., 2006b). The TTS server is prepared for supporting of the SSML language in the future.

### 3.2.4 Dialogue manager server

The main module in the Slovak SDS is the dialogue manager server (Ondáš & Juhár, 2005), which controls the interaction between system and user and it is responsible for generating requests to the system's actions like playing prompts, recognizing user's utterance, obtaining information from the Internet and etc. De facto it also manages other servers of the SDS. The dialogue manager module in the Slovak SDS fully supports VoiceXML 1.0 specification and significant part of VoiceXML 2.0 specification. Therefore the heart of the manager is an interpreter of VoiceXML mark-ups. The principle scheme of DM module is shown on Fig. 3. The core of the manager is the aforementioned VoiceXML interpreter together with ECMAScript engine and XML Parser, which directly relates to interpretation of VoiceXML scripts. Output generator is responsible for implementation of the Prompt selection algorithm and its output can have format of SSML document. The input processor together with the grammar manager constitutes the NLU subpart of the dialogue manager. Grammar manager integrates the Grammar activation algorithm as well as semantic

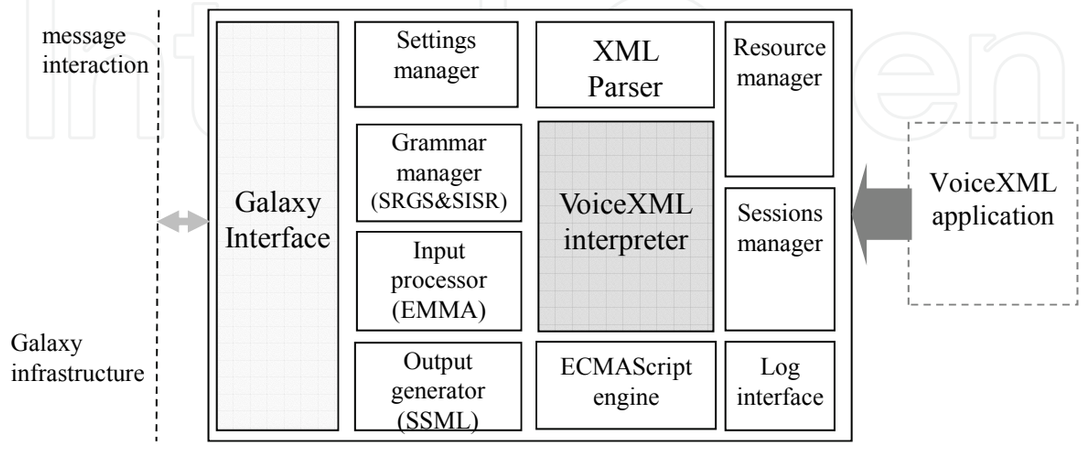



Fig. 3. The principle scheme of the Dialogue manager module

interpretation according SISR markups. The input processor generates requirements on acquiring user's input (DTMF or spoken utterance) and initializes processing of incoming inputs. Resource manager implements the Resource fetching algorithm. Settings manager is responsible for management of actual settings of the DM as well as SDS and for the realization of *property* markup of the VoiceXML language. Sessions manager creates, manages and destroys the user's session. Recording of information about interaction and internal state of DM is done by Log interface. The last part of the manager is the Galaxy interface, which connects it with remaining part of the system.
A dialog with the user starts after the message from the I/O server about the incoming call. Then DM initializes a new session and loads the VoiceXML script (application), interprets it and according to those it generates requirements on services (functions) of others system modules, in such manner, that the dialogue with the user is led.

### 3.2.5 Information server

Next part of the Slovak spoken dialogue system is the information server or backend server. It is capable of retrieving the information contained on the suitable web-pages according to the dialogue manager requests, extracting the requested data, analyzing them and if they are considered being valid, it returns the data to the DM. These tasks are performed by unique web-wrappers, which are the important parts of the server. The web-wrapper is responsible for the navigation through the web-server, data extraction from the web-pages and their mapping in a structured format (XML), convenient for further processing. In most cases the wrapper is specially designed for one source of data; thus for combining the data from different sources, several wrappers must be designed. Wrappers are designed to be as robust against changes in the web-pages structure as possible. To speed up the system (to eliminate the influence of long reaction times of the www-pages) and to assure drop-out resistance with simultaneous refresh of the information, automatic periodic download and caching of the web-pages content were introduced. The server is open to future applications by the possibility of creating web-wrappers for new services and adding them to the existing wrappers for Weather Forecast and Train Timetable services.

### 3.2.6 Evaluation server

The evaluation server tracks communication among servers and computes a set of interaction parameters. Möller in (Möller, 2005) defines interaction parameters as measures related to the system and the service which can be collected during the interaction. They enable quantifying parameters of the system and services and according to them concluding the quality of the evaluated system and services. Several interaction parameters can be obtained without human expert that leads to the automatic evaluation. This aspect is very important because of reducing costs spending on the evaluation during the system's lifecycle. Evaluation server in the Slovak SDS enables collecting the interaction parameters like dialogue duration, system prompt delay, user response delay, average number of noinput or nomatch events and so on. It produces several log files with those parameters for each channel and calculates overall as well as channel statistics.

## 3.3 Voice services provided by Slovak SDS

The Slovak SDS provides two pilot voice services – Weather forecast service for Slovakia and Timetable services (for Slovak Railways, Buses and City buses of Košice). The services

have been designed as a system-directed with the open structure based on subdialogues (Ondáš, 2007). The data are retrieved directly from the Internet. Provided services enable metacommunication in a form of a set of keywords like help, back, repeat, etc.
The Weather forecast service provides information about weather for thirty towns in Slovakia up to three days ahead.
The Timetable services allow finding out the connection between more than ten thousand points (stations or POIs) in Slovakia. After the welcome message and the service selection subdialog the user is prompted to provide information about start point, end point, date and time of the departure. In the designed dialog we have used explicit conditioned confirmation after a pair of input items. Using of the implicit confirmation require more sophisticated speech grammars, therefore we did not select this form of confirmation. Conditioned confirmation means that the input items collected from the user are confirmed only when their confidence level is lower than chosen threshold. After the confirmation of all provided information from the user, the system generates a query to retrieve data from internet, in which all items are sending to the web server. Then the obtained data are played to the user. Playing of the information is divided in to two layers. In the higher layer only the basic information is played. If the user wants to play detailed information, he can enter the lower layer, in which detailed information will be played.
The system is in public experimental running since January 2006. It runs in multi-user mode through PSTN and GSM telephony network and through VoIP (Skype&SIP). Until now more than 8000 calls were answered by the Slovak SDS.

## 4. Subjective evaluation of the Slovak SDS

Spoken dialogue systems (SDS) are nowadays widely used in several domains. This fact brings a need of evaluation, comparison and categorization of dialogue systems and their services. The quality of the interaction with a telephone-based speech service can be addressed from two separate points-of view. System developers are concerned about system/system's modules performance. From the user's point-of view, the perceived quality or overall opinion are the most important aspects (Möller, 2005). Using subjective measures is the only way of finding out user's opinion of the system. Objective measures, such as performance, do not have a direct attachment to the user satisfaction (Hartikainen et al., 2004). From that point of view, measurement of interaction parameters can bring only some information about system behaviour and properties. While extensive effort has been put to the definition and measurement of efficiency and effectiveness in user interactions with speech systems (ITU-T P.851, 2003), comparatively little emphasis has been put on measurement of subjective satisfaction (Hone & Graham, 2001). There are few projects, which are focused on this domain. The first one is the SASSI (Subjective Assesment of Speech-System Interface) methodology, which considers the "validity" and "reliability" aspects of the questionnaires as the most important. The questionnaires in this project are prepared in iterative design process, in which, at first, a pool of attitude statements is designed and in several iterations only a relevant set of statements is selected. An initial 50 item questionnaire was designed. Each attitude statement was rated according to seven points scale. Authors identify six factors or quality aspects: perceived system response accuracy, likeability, cognitive demand, annoyance, habitability and speed (Hone & Graham, 2001).
SERVQUAL method for subjective quality evaluation was adopted from the area of marketing applications and it is suitable also for subjective evaluation of dialogue systems.

Authors in (Hartikainen et al., 2004) view the spoken dialogue system as a service, which is provided to the users. This method is based on two principles: Service quality can be divided into dimensions, and measured as a difference of expectations and perceptions (Parasuraman, 1988). SERVQUAL method defines five service quality dimensions: tangibles, reliability, responsiveness, assurance and empathy. SERVQUAL method uses questionnaires by 22 items. Respondents assess how the reality meets their expectations.
The subjective evaluation methods providing information about the quality of telephone services are also described in the ITU Recommendation P.851 (ITU-T P.851, 2003). The evaluation methods described in this recommendation address different aspects of quality from a user's point of view, as are the usability of the service, the communication efficiency, task and service efficiency, user satisfaction, perceived speech input and output quality, the system's cooperability, etc (Möller, 2005). Described methods are based on laboratory experiments in which subjects interact with the spoken dialogue system in order to perform a pre-defined, realistic task. Then they fill a set of questionnaires, which reflected their opinion on assessed system and services. Recommendation contains also the set of questions (items) related to described aspects of quality and the examples of test scenarios.
Within years 2006 to 2008 we have proposed a new subjective evaluation method, which adopts methodology described in ITU P.851 Recommendation. There were several motivations for proposing a new (modified) method. The first, perceived quality of services provided by the Slovak dialogue system has never been adequately evaluated. Second motivation was that the methods introduced above cannot be simply used for evaluation of the Slovak SDS, because they do not take in to consideration usage of the W3C languages and Galaxy infrastructure. Our goal was also to propose the simple method with comparable and understandable result in a form of school grading system. The reasons for the selection of the ITU P.851 Rec. as a basis for the new method are summarized in (Ondáš et al., 2008).
The designed method is questionnaire-based and it is appointed to realization of subjective evaluation experiments for obtaining the user judgments and opinions of the spoken dialogue system and service quality. Within this method a set of questionnaires, test scenarios and rating scales are defined. Also there are defined both, new categorization of quality aspects and grades of quality for rating of these categories.

### 4.1 Evaluation questionnaires and methodology

The questionnaire form of evaluation was selected as the most suitable form of obtaining information about perceived quality of dialogue system and its services. From a set of questions/items, defined in ITU T. Rec P.851, three types of questionnaires were prepared:

- Questionnaire A - contains items related to user's background, their knowledge about domain and system. The items in this questionnaire were modified for testing the Slovak timetable information service. It contains 12 items.
- Questionnaire B - comprises 17 items related to individual interaction with the system.
- Questionnaire C - contains 14 closed and 3 open items related to the overall impression of the system and provided services.

Tab. 1. contains item numbers according to ITU T.Rec P.851, which was adopted for the designed questionnaires. For obtaining a complex view on quality aspects it is necessary to interact with the system more than once. Also the sufficient motivation on the side of the test subjects is required. Because of these facts, a set of test scenarios was designed. Each

user (test subject) makes two calls on the system's telephone number. First, the test subject fills in the questionnaire A. Then they make a call on the one of the telephone numbers of the Slovak SDS and in spoken interaction realizes given "common scenario". After the interaction they fill in the questionnaire B (B1), in which they assess the prior interaction. Then they make a second call and carry out the given "individual scenario". They assess the prior interaction in questionnaire B (B2). At last the test subject fills in the questionnaire C, in which they assess the both interactions and their overall impression of the system and its service.

| Type of questionnaire | Item numbers |
|---|---|
| Type B | overall impression, 1, 2, 4-9, 11, 15 - 17, 19 - 22 |
| Type C | 1, 3, 4, 6 - 8, 10, 11, 13 - 21 |

Table 1. Items used for building of questionnaires B and C

### 4.2 Processing the results

The questionnaires processing consists of two operations - coding and categorization. The coding is a substitution of data by symbols, which will be used in statistical methods. The rating scale, which can hold values from 0 to 1, was designed for coding of items of the questionnaire. The "0" value represents the worst (the lowest) level of the quality (property). Conversely, the "1" value represents the best (the highest) level of the quality.
The items/questions in the questionnaires were coded with five types of scales:

- the growing 7-level scale
- the decreasing 7-level scale
- 5-level Likert scale
- backward 5-level Likert scale
- 5-level centered scale (the highest value in the middle)

There was designed categorization on six aspects of quality (categories). The first four categories are the same as in ITU T. recommendation. The last two categories were created for obtaining direct information about user's satisfaction and about usability of the system and its services. The designed categories are the following:

- *information obtained from the user; communication with the system; system's behaviour; dialog; user satisfaction; usability*

Each category is characterized by the set of questions from questionnaires B (1 and 2) and C. One question can be assigned to several categories. Then the categories are rated by one of the quality grades. There were six quality grades designed, from A to FX according to standard school grading system. For each category the quality grade is evaluated as an arithmetic mean of all item's responses for the given category (their score, assigned during coding) in percentage.

### 4.3 Evaluation experiments

Two evaluation experiments were performed with the basic setup of the Slovak spoken dialogue system. The first experiment was carried out with 26 test subjects (students). They made 52 interactions with the system. They filled in 104 questionnaires of A, B and C type. The calls were made through PSTN network in two acoustic environments - in the office (the

silent environment) and in the laboratory with twelve students (noisy environment). Evaluation was performed on Slovak railway timetable service. The second column in Table 2 contains the results of the experiment. Experiment is in more detail described in (Ondáš et al., 2008).

| Category | Experiment 1 | Experiment 2 |
|---|---|---|
| Information obtained from the system | C (79.4%) | *C (79.5%)* |
| Communication with the system | D (65.9%) | C (75.1%) |
| Behaviour of the system | C (78.5%) | C (76.6%) |
| Dialog | C (71%) | D (67.9%) |
| User satisfaction | D (64.5%) | C (71.6%) |
| Usability | D (62.3%) | D (63.2%) |

Table 2. Results of evaluation experiments

The second experiment was carried out with 24 test subjects. Evaluation methodology was reduced because of several facts. The most important fact was that all test subjects had some experience with evaluated SDS. Therefore, they did not fulfil questionnaire type A and they interacted with the system only once. Our experiences have shown that there is a high correlation between items in questionnaire B and C. Therefore, test subjects fulfilled only questionnaire type C and classification of the items was modified by mapping questions from that questionnaire. The proposed reduction in evaluation methodology led to absence of items for the first category. The value in this category was replaced by the interaction parameter *successfulness,* defined in (Ondáš & Juhár, 2007), which was obtained by expert evaluator. The experiment and evaluation was performed on City bus timetable service. The third column in the Table 2 contains results of the second experiment. The second experiment is in more detail described in (Ondáš et al., 2009).

## 5. Conclusion

The W3C Speech Interface Framework specifications have become industrial standards in domain of voice user interfaces, voice browsers as well as spoken dialogue systems. They enable rapid development of that systems and a large range of voice applications. The main advantages of the Speech Interface Framework consist of portability, uniformity of design process, easiness of designing the new systems and services and the possibility of automatic generation of such applications. A set of XML-based languages provides a good starting point also for integration with web applications.

Proposed article shows the possibility of combining free resources with up to date standards. Relatively old Galaxy infrastructure, used in the Slovak SDS, provides surprisingly great solution for integrating the W3C Speech Interface Framework languages, with the properties, that can lead to building the complete system for real environment. Of course, the success of such system relies on technologies employed in system servers (speech recognition, text-to-speech, etc).

In the future our work will be focused on supporting all SIF languages by our dialogue system as well as on enabling more natural interaction (advanced dialogue) between system

and user. Nowadays the interaction in the dialogue system has a form of filling in the questionnaire by voice. That means, the system ask the user for providing information, which determines its request and then it tries to obtain the answer from the web localities and deliver it to the user by speech synthesis. It is clear, that the nature of conversation is an interactive process rather than a structural product (Jokinen, 2009). The impossibility of asking the system by user is the main disadvantage of the VoiceXML language. The user mainly answers to the system questions. Of course, the answer of the user can be in a form of question, for example: "Could you help me?", but there do not exist the mechanism for real "user initiative". The VoiceXML language also allows so-called "interaction with mixed initiative", but this mode only enables user to provide several information in one utterance. In our next work, we plan to focus on the solution for enabling interaction with user intitiative in the range of Speech Interface Framework.

## 6. Acknowledgements

The work presented in this paper was supported by Slovak Research and Development Agency under research projects APVV-0369-07 and VMSP-P-0004-09 and Ministry of Education of Slovak Republic under research projects VEGA-1/0065/10.

## 7. References

Delgado, López-Cózar R. & Araki M. (2005). Spoken, multilingual and multimodal dialogue systems: development and assessment. ISBN 0470021551, 9780470021552, John Wiley, 2005, Michigan university, 261 p., 2005

Hartikainen, M. & Salonen, E. & Turunen, M. (2004). Subjective evaluation of spoken dialogue systems using SERQUAL method. In ICSLP 2004. International conference on spoken language processing: proceedings. ICSLP, 2004

Hone, Kate S. & Graham, R. (2001). Subjective assessment of speech-system interface usability, In Eurospeech 2001, pp. 2083-2086, Scandinavia 7th European Conference on Speech Communication and Technology, Aalborg, Denmark, September 3-7, 2001 ed. by Paul Dalsgaard, Borge Lindberg, Henrik Benner, and Zheng-hua Tan; ISCA Archive, 2001

ITU-T Rec. P.851, Subjective Quality Evaluation of Telephone Services Based on Spoken Dialogue Systems, International Telecommunication Union, Geneva, 2003. http://www.itu.int/rec/T-REC-P.851-200311-I/en

Jokinen, K. (2009). Constructive dialogue modelling: speech interaction and rational agents, Wiley series in agent technology, ISBN 0470060263, 9780470060261, John Wiley and Sons, 2009, 160 p., 2009

Juhár, J. & Ondáš, S. & Čižmár, A. & Jarina, R. & Rusko, M. & Rozinaj, G. (2006). Development of Slovak GALAXY/VoiceXML Based Spoken Language Dialogue System to Retrieve Information from the Internet, Proceedings of Interspeech 2006, Pittsburg, USA, Sept. 17-21, 2006, paper 2056-Mon2FoP.10, 2006, Pittsburg

Juhár, J. et al.: Voice Operated Information System in Slovak, Computing and Informatics, Vol. 26, 2007, 577 - 603

Lihan, S. & Juhár, J. & Čižmár A. (2005). Crosslingual and Bilingual Speech Recognition with Slovak and Czech SpeechDat-E Databases, In Proc. Interspeech 2005, Lisabon, Portugal, September 2005, pp. 225 – 228, 2005

MAVBS, (1999). Model Architecture for Voice Browser Systems, W3C Working Draft: http://www.w3.org/TR/voice-architecture/, December 1999.

McTear, F., M. (2005). Spoken Dialogue Technology: Toward the Conversational User Interface, ISBN 1-85233-672-2, Springer-Verlag London Limited 2004, United States of America, 2004

MIT (2008). http://groups.csail.mit.edu/sls//technologies/galaxy.shtml, 2008

MIWG, (2010). Multimodal Interaction Working Group website: http://www.w3.org/2002/mmi/

Möller, S. (2005). Evaluating telephone-based interactive systems, In ASIDE-2005, Aalborg, Denmark, November 10-11. 2005, paper 42

Ondáš, S. & Juhár, J. (2005). Dialog manager based on the VoiceXML interpreter. In: Proc. of the DSP-MCOM 2005: The 6th international conference on Digital Signal Processing and Multimedia Communications, ISBN 80-8073-313-9, pp. 80-83, September 13-14, 2005, Košice, Slovak Republic, Technical university of Košice, Košice, 2005

Ondáš, S. (2007). Principles of voice services design for IRKR communicator, In: 7th PhD Student Conference and Scientific and Technical Competition of Students of Faculty of Electrical Engineering and Informatics Technical University of Košice: Proceeding from conference and competition, pp. 17-18, ISBN 978-80-8073-803-7, FEI TU Košice, May 2007, Košice

Ondáš, S. & Juhár, J. (2007). Automatic evaluation of Slovak spoken language dialogue system, In: ECMS 2007 & Doctoral School: 8th international workshop on Electronics, Control, Modelling, Measurement and Signals, pp. 91-96, ISBN 978-80-7372-218-0, Liberec, Czech Republic, May 21-23, 2007, Technical University of Liberec, 2007

Ondáš, S. & Juhár, J. & Čižmár, A. (2008). Evaluation of the Slovak spoken dialogue system based on ITU-T, In: Text, Speach and Dialogue: 11th International conference, TSD 2008, Brno, Czech Republic, September 8-12, 2008, Berlin, Springer-Verlag, ISBN 978-3-540-87390-7, ISSN 0302-9743, pp. 633-640, 2008

Ondáš, S. & Juhár, J. & Pleva, M. (2009). The city bus timtable voice service for Kosice, In: SPA 2009: signal processing: algorithms, architectures arrangements, and applications, ISBN 978-83-62065-00-4, pp. 154-158, Poznan, University of Technology, Poland, 2009

Parasuraman, A. & Zeithaml, V.A. & Berry, L.L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality, Journal of Retailing, 64, 1, 1988.

Pleva, M. et al. (2008). Concept of spoken dialogue system based on voice over IP telephony. In: RTT 2008: Research in Telecommunication Technology 2008, 9th International Conference, Bratislava, STU, 2008, ISBN 978-80-227-2939-0, pp 3, 2008

Polifroni, J. & Seneff, S. (2000). GALAXY-II as an Architecture for Spoken Dialogue Evaluation, Proceedings of Second International Conference on Language Resources and Evaluation (LREC), Greece, May 31-June 2, 2000, Athens

Pollak, P. et. al. (2000). SpeechDat(E) - Eastern European Telephone Speech Databases, In Proceedings LREC 2000 Satellite workshop XLDB - Very large Telephone Speech Databases, pp. 20-25, Athens, Greece, May 2000, Athens

Rosenberg, J. et al. (2006). SIP: Session Initialization Protocol. June 2002-2006, IETF, http://datatracker.ietf.org/doc/rfc3261/

Rusko, M. & Trnka, M. & Darjaa, S. (2006)a. MobilDat-SK - A Mobile Telephone Extension to the SpeechDat-E SK Telephone Speech Database in Slovak, In press: SPEECOM 2006, Sankt Peterburg, Russia, July 2006, Sankt Peterburg

Rusko, M. & Trnka, M. & Daržagín, S. (2006)b. Three Generations of Speech Synthesis Systems in Slovakia, In: Proc. of XI International Conference Speech and Computer, SPECOM 2006, Sankt Peterburg, Russia, 2006. ISBN 5-7452-0074-X, pp. 297-302.

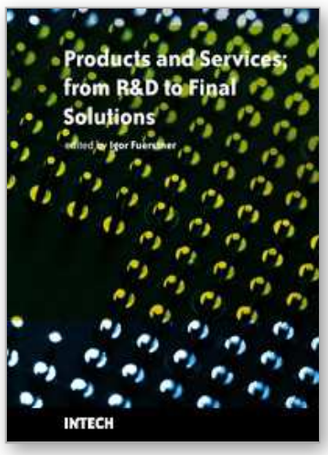
VBA, (2010). Voice Browser Activity website: http://www.w3.org/Voice/

VXMLForum, (2010). VoiceXML Forum website: http://www.voicexml.org/

Walker, A. M. & Litman, J. D. & Kamm, A. C. & Abella, A. (1997). PARADISE: A Framework for Evaluating Spoken Dialogue Agents, in Proceedings of ACL/EACL 35th Annual Meeting of the Association for Computational Linguistics, San Francisco: Morgan Kaufmann, 1997, pp. 271-280.

W3C, (2010). World Wide Web Consortium website: http://www.w3.org/

Young, S. (2004). ATK: An application Toolkit for HTK, version 1.3", Cambridge University, January 2004




**Products and Services; from R&D to Final Solutions**
Edited by Igor Fuerstner

ISBN 978-953-307-211-1
Hard cover, 422 pages
**Publisher** Sciyo
**Published online** 02, November, 2010
**Published in print edition** November, 2010

Todayâ€™s global economy offers more opportunities, but is also more complex and competitive than ever before. This fact leads to a wide range of research activity in different fields of interest, especially in the so-called high-tech sectors. This book is a result of widespread research and development activity from many researchers worldwide, covering the aspects of development activities in general, as well as various aspects of the practical application of knowledge.

**How to reference**
In order to correctly reference this scholarly work, feel free to copy and paste the following:

Jozef Juhár and Stanislav Ondáš (2010). Development and Evaluation of the Spoken Dialogue System Based on W3C Recommendations, Products and Services; from R&D to Final Solutions, Igor Fuerstner (Ed.), ISBN: 978-953-307-211-1, InTech, Available from: http://www.intechopen.com/books/products-and-services--from-r-d-to-final-solutions/development-and-evaluation-of-the-spoken-dialogue-system-based-on-w3c-recommendations

INTECH
open science | open minds

**InTech Europe**
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

**InTech China**
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen