MOBILTEL - Mobile Multimodal Telecommunications dialogue system based on VoIP telephony

Anton Čižmár, Matúš Pleva, Ján Papaj, Ľubomír Doboš, Jozef Juhár

Dept. of Electronics and Multimedia Communications,
Technical University of Košice,
Letná 9, 042 00 Košice, Slovak Republic,
Email: Anton.Cizmar@tuke.sk, Matus.Pleva@tuke.sk, Jan.Papaj@tuke.sk,
Lubomir.Dobos@tuke.sk, Jozef.Juhar@tuke.sk

Abstract *– In this paper the project MobilTel is presented. The communication itself is becoming a multimodal interactive process. The MobilTel project provides research and development activities in multimodal interfaces area. The result is a functional architecture for mobile multimodal telecommunication system running on handheld device. The MobilTel communicator is a multimodal Slovak speech and graphical interface with integrated VoIP client. The other possible modalities are pen – touch screen interaction, keyboard, and display on which the information is more user friendly presented (icons, emoticons, etc.), and provides hyperlink and scrolling menu availability.*

We describe the method of interaction between mobile terminal (PDA) and MobilTel multimodal PC communicator over a VoIP WLAN connection based on SIP protocol. We also present the graphical examples of services that enable users to obtain information about weather or information about train connection between two train stations.

Keywords: *multimodal services, PDA, voice over IP, speech recognition.*

I. INTRODUCTION

The main goal of this project is the research and development in the area of mobile multimodal telecommunication systems, which allows access to information from different areas through mobile multimodal terminal and human - machine interaction with natural speech, with support of another - mainly graphical - modalities.

The solution of the project is furthermore the goal of information exchange and acquisition of new knowledge from the area of the research, development and use of mobile telecommunication systems and services, automatic speech recognition, speech synthesis, automatic speech and multimodal dialog systems, network programming, voice over IP protocols, and other subjects according to the solving of the point of this project.

MobilTel Communicator enables multimodal multi-user interaction in Slovak language through telecommunication or IP network to find information in databases or Internet websites. The speech, graphical user interface, pen and keyboard on PDA device are on of the possible and also available modalities.

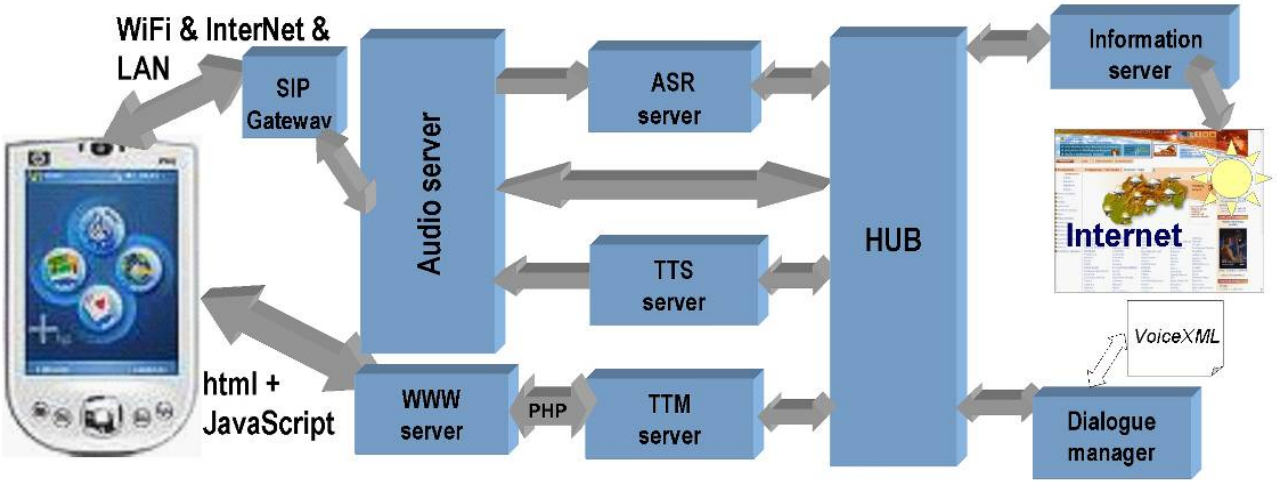

Fig. 1. MobilTel Communicator architecture

II. MULTIMODAL ARCHITECTURE

The MobilTel communicator is based on a distributed 'hub-and-spoke' Galaxy architecture used in DARPA Communicator [12], [13].

Each module (server) seeks services from the HUB and provides services to the other modules through the Galaxy HUB.

It's based on easy Plug & Play approach, which means that plugging new servers is very easy and handful thanks to unified frame-based message communication with the HUB. All communication goes trough the HUB, which can distribute the messages to more then one recipients and can also change the message or the structure and make a log files.

A. Servers of the communicator

ASR server – There are two isolated words recognizers with thousands words capacity:
1.The ATK (An Application Toolkit for HTK) based ASR module [3]
2.Sphinx-4 based ASR module [2], [10].
- Acoustic models are context dependent (triphone), trained in a training procedure compatible with "refrec" [1], [4] trained on SpeechDat-SK [5], [6] and MobilDat-SK [7] databases.

HUB server - developed in DARPA Communicator project [15], programmable by script files called "hub programs".

Information (Backend) server - capable of retrieving the information from the web sites according to the Dialogue Manager (DM) requests, extracting the specified data and returning them in the XML format to the HUB.

TTS server - diphone synthesizer, based on concatenation of diphones with Time Domain Pitch Synchronous Overlap and Add (TD-PSOLA) similar algorithm [8].

Dialogue manager - Based on VoiceXML 1.0 interpreter, fundamental components are: XML Parser, VoiceXML interpreter and ECMAScript unit The dialog manager needs an external message wrapper to the HUB [11].

Telephony server - Connects the whole system to the telecommunication or VoIP network, supports telephone hardware card - Dialogic D120/41JCT-LSEuro, sound card or VoIP (SIP) connection. This server needs a direct real-time data connection with ASR and TTS server to transmit speech data.

WWW server –Apache free web server which provides the GUI interface to the PDA device. The source is PHP documents, which represents the dialog according to the VocieXML document. The optional Javascript can provide additional features to the GUI interaction (interactive calendar for date input, just in time filter for city input, etc.)

TTM (Text To Multimodality) server – provides next modalities, trough direct access to HUB with messages generated from files, and also logging messages from HUB to another files. These files are monitored real-time, and any other program or script can interact with the dialog of the communicator. In this case the PHP script running on the webserver monitors and generates these files and provides GUI interface to the PDA, with possibility to interact also by clicking on hyperlink, choosing from the scrolling menu or writing the option with the keyboard. Also the required information is user friendly graphically visualized (moving icons, pictograms, etc.).

III. VoIP INTEGRATION USING SIP

The Session Initiation Protocol (SIP) is a signaling protocol for call setup and teardown including video, voice and instant messaging [1]. This protocol is the most frequently used standard between cheap VoIP providers.

There are also a lot of complete open source clients for many platforms. The SIP protocol is transport-independent, because SIP can be used with UDP, TCP, SCTP, etc. The protocol is text-based (built on principles of HTTP and SMTP), allowing for humans to read and analyze SIP messages.

The problem is NAT (network address translation) traversal and firewalls. SIP uses a new UDP connection for every call. There are o lots of possibilities for solving the problem: Universal Plug and Play (UPnP), STUN (Simple Traversal of UDP through NATs), Connection Oriented Media (Comedia), RTP Relay (TURN), Application Layer Gateway (ALG) or different tunneling protocols. Anyway this is the weakness of SIP protocol.

Then a VoIP (SIP) gateway (with 2 FSX ports) was also purchased for testing SIP clients from various platforms and the open source client chosen for integration to Galaxy-based AudioServer.

The chosen open source VoIP architecture was pj-sip [18]. It's a complete SIP protocol stack with high performance and extremely portable source code. This small VoIP client was tested on Windows, Windows Mobile (for calls from PDA devices) and Linux platform.

The next step is the VoIP client integration to Galaxy based AudioServer, which is responsible for audio data exchange between the spoken dialog server (see Fig. 2 below) and the SIP provider.

The audio data are transferred using a small memory buffers and the codec used during the SIP transmission could be chosen from available codec list, but both clients needs to agree on chosen codec before the audio transmission starts. Every SIP client needs a registration to some VoIP provider.

IV. EXAMPLES OF MULTIMODAL SERVICES

There are two testing multimodal services: "Railway scheduler" and "Weather forecast" for testing of MobilTel architecture communicator. In these services is a combination of speech modality and other modalities. On first screen we can choose interactively clicking on the graphical representation of these two services: "Weather" and "Railway scheduler" (see Fig.3).

During the interaction the global choices could be chosen from the above menu: Home, Back, Help, Mute (to turn on/off the loudspeaker – if we do not need to hear the speech from TTS server – we are in noisy environment).

A. Railway scheduler service

Railway scheduler provides information about train connections. We can select place from and where we want to travel, the date and time (default is the present time and today) when we want to travel (see Fig.3).

Selecting all inputs we can get information about the closest train connection, price, name and type of the first connection train. If there is some changing station there will be all transfer trains presented below. Also the next possible connection button will provide the information about the next train. Information displayed on the dialog screens and the result screen is presented below.

Fig. 3. Examples of service GUI for Railway scheduler

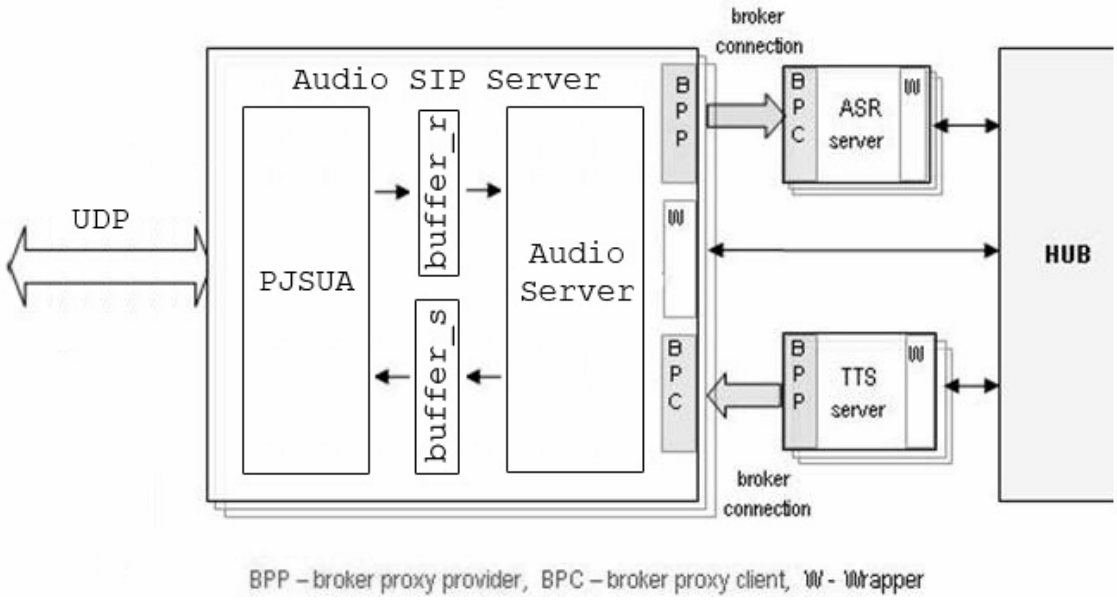

Fig. 2. MobilTel Audioserver based on SIP architecture

B. *Weather forecast service*

Weather service provides information about weather forecast in the district towns of Slovakia [9]. We select the city from scrolling menu and the day (see Fig. 4). We can also request the current weather information from the last hour.

The result for weather forecast is the maximal and minimal daily temperature and the weather state (rainy, sunny, etc.) presented like pictogram.

The actual weather information provides only current temperature during observed during the past hour and also the air pressure. The time of the temperature observation is also presented.

Fig. 4. Examples of service GUI for Weather forecast

V. CONCLUSIONS

The MobilTel communicator is the first functional multimodal user interface in Slovak language for mobile devices. The next project activities will be oriented to optimizing and testing of the system. We will try to select optimal solution and increase the robustness by optimization of duplex voice transmission between PDA and multimodal server, choosing an optimal codec for voice recognition and transmission over wireless LAN, optimization of GUI on the server and PDA side. Next, we will evaluate the design process of creation the graphical and speech dialog together, for better fusion of the dialog based on VoiceXML and PHP. Then the optimizing of the mixed initiative dialog management, for better resolution of the user's choice made using more modalities (touch screen and voice) will be realized. Next part of the project will be redesign and testing of the multimodal telecommunication services and optimization of multimodal dialog for services "Weather" and "Railway scheduler".

In the final phase the multimodal version of MobilTel communicator system will be tested by freeing the communication for PDA devices connected to the laboratory WiFi AP (access point) and publishing the needed connection information to the students on the project website [16].

ACKNOWLEDGEMENTS

Research described in the paper was supported by the Slovak Research and Development Agency under research projects APVT-20-029004 and APVV-0369-07 and FP7 EU project INDECT No. 218086.

REFERENCES

[1] B. Lindberg, F. T. Johansen, N. Warakagoda, G. Lehtinen, Z. Kačič, A. Žgank, K. Elenius, G. Salvi, "*A noise robust multilingual reference recognizer based on SpeechDat (II)*", in Proc. ICSLP 2000, Beijing, China, 2000.

[2] P. Lamere, P. Kwok, W. Walker, E. Gouvea, R. Singh, B. Raj, P. Wolf, "*Design of the CMU Sphinx-4 decoder*", in Proc. Eurospeech 2003, Geneve, Switzerland, September 2003, pp. 1181–1184

[3] S. Young, "*ATK: An application Toolkit for HTK, version 1.3*", Cambridge University, January 2004.

[4] S. Lihan, J. Juhár, A. Čižmár, "*Crosslingual and Bilingual Speech Recognition with Slovak and Czech SpeechDat-E Databases*", In Proc. Interspeech 2005, Lisabon, Portugal, September 2005, pp. 225 – 228.

[5] R. Winski, "*Definition of corpus, scripts and standards for fixed networks*", Technical report, SpeechDat-II, Deliverable SD1.1.1., workpackage WP1, January 1997.

[6] P. Pollak, J. Cernocky, J. Boudy, K. Choukri, M. Rusko, M. Trnka, et al., "*SpeechDat(E) „Eastern European Telephone Speech Databases*", in Proc. LREC 2000 Satellite workshop XLDB - Very large Telephone Speech Databases, Athens, Greece, May 2000, pp. 20-25.

[7] M. Rusko, M. Trnka, S. Darjaa, "*MobilDat-SK - A Mobile Telephone Extension to the SpeechDat-E SK Telephone Speech Database in Slovak*", SPEECOM 2006, Sankt Peterburg, Russia, July 2006.

[8] S. Darjaa, M. Rusko, M. Trnka, "*Three Generations of Speech Synthesis Systems in Slovakia*", SPEECOM 2006, Sankt Peterburg, Russia, July 2006.

[9] I. Gladišová, Ľ. Doboš, J. Juhár, S. Ondáš, "*Dialog Design for Telephone Based Meteorological Information System*", in Proc. DSP-MCOM 2005, Košice, Slovakia, Sept., 2005, pp. 151-154.

[10] M. Mirilovič, S. Lihan, J. Juhár, A. Čižmár, "*Slovak speech recognition based on Sphinx-4 and SpeechDat-SK*", in Proc. DSP-MCOM 2005, Košice, Slovakia, Sept. 2005, pp. 76-79.

[11] S. Ondáš, J. Juhár, "*Dialogue manager based on the VoiceXML interpreter*", in Proc. DSP-MCOM 2005, Košice, Slovakia, Sept. 2005, pp.80-83.

[12] http://www.sls.csail.mit.edu/sls/technologies/galaxy.shtml

[13] http://communicator.sourceforge.net/

[14] http://fife.speech.cs.cmu.edu/Communicator/index.html

[15] http://communicator.colorado.edu/

[16] http://mobiltel.tuke.sk/

[17] http://www.pjsip.org/