ASR & TTS Applications in Western Balkan Countries

Vlado Delić, *Member, IEEE*

***Abstract* — Automatic speech recognition and text-to-speech conversion are language dependent technologies, developed for a small number of languages. ASR and TTS for Serbian have achieved a sufficient level of quality and their first applications have been launched. A brief review of the applications in West Balkan countries is given in this paper. They are mostly dedicated to the persons with disabilities, but several other ASR&TTS applications have appeared as well.**

***Key words* — Automatic speech recognition (ASR), computer-telephony integration (CTI), persons with disabilities, text-to-speech conversion (TTS).**

I. Introduction

SPEECH technologies R&D for a certain language requires knowledge in areas such as linguistics, phonetics, acoustics, mathematics, programming as well as digital signal processing [1]. It is necessary to bring together the knowledge from these areas and implement it into the available computer resources to enable a computer to understand human speech – using automatic speech recognition (ASR), as well as to respond via speech – using text-to-speech synthesis (TTS). Both are very complex multidisciplinary problems requiring engagement of teams of experts from aforementioned areas, as well as ample time and financial resources. For that reason, solutions for widely spoken languages were the first to appear. However, due to the aforesaid language dependency, they could not have been used in areas where other languages were spoken.

Once humans are able to address a computer in their native language and a computer is able to respond in the same language, it will be possible to "talk" to other appliances, as well as industry machines, cars, toys and robots, or to a remote computer via telephone. We will have an increasing need to speak to appliances in our midst and thus it is extraordinarily important to develop speech technologies, each of us for their own language. There is a profusion of languages in the world and each nation strives to preserve and protect their language. Speech technologies have the potential to overcome an evident language barrier between collocutors who do not speak the same language. We have succeeded in overcoming certain research barriers related to the complex morphological structure of Serbian and kindred South Slavic languages [2-10]. In this way further progress in development of speech technologies for the Serbian language has been made possible and we can now proudly state that we have introduced our language into a relatively narrow circle of world languages for which speech technologies have attained the necessary level of quality and that they find their first wide applications.

The paper will present first applications in Western Balkan Countries (WBC): Serbia, Croatia, Bosnia and Herzegovina, Macedonia and Montenegro [2-3]. WBC include Albania as well, however the Albanian language does not belong to the Slavic group.

Speech technologies are still far from being perfect and will remain so for many years to come, unable to interpret human speech as correctly as a human does or produce speech that would be indistinguishable from human speech. However, the author of this paper has already had the pleasure to hear a computer pronounce certain phrases faster and more precisely than any human can. The same computer is already capable of talking to tens of people simultaneously, understand speech commands issued by humans and talk back using synthesised speech. This synthesised speech may not be as rich in prosody features as original human speech, but we already understand it much better than, e.g., speech produced by many persons with speech disabilities. This shows that at least persons unable to speak can make use of a TTS system to say what they want using synthesised speech. Quality of synthesised speech has been significantly improved in the last several years and it is being used by the visually impaired to read books, newspapers and letters unaided, using only a speech enabled computer. In this way the visually impaired are more equal in their education, access to information as well as privacy in written communication.

On the other hand, recognition of spontaneous speech with words from a large vocabulary (more than several thousand words) has still not been developed for a majority of world languages. However, small and medium-sized vocabulary speech recognition (several tens or hundreds of words) can be of quite satisfactory use in a guided dialogue between humans and machines. For instance, a cleverly designed and guided dialogue between a human and a computer (e.g. via phone) can use a different set of words in each phase, words that are easy to distinguish. In the next phase a new set of words – a new vocabulary – can be used, making the human-machine dialogue very rich and

This work was supported in part by the Ministry of Science of Serbia within the Project "Development of speech technologies in Serbian and their application in 'Telekom Srbija'" (TR-6144A).

Vlado D. Delić, Fakultet tehničkih nauka, Trg D. Obradovića 6, Novi Sad, Srbija (telefon: +381-21-485-2533; e-mail: vdelic@uns.ns.ac.yu).

efficient. The most important thing for an application designer is to be aware of the actual capabilities of the technology, not to rely on theoretical data on error rates only, but put the interactive voice response (IVR) application to a proper test in real conditions.

Computer-telephony integration (CTI) applications of ASR and TTS should make most of the advantages of speech technologies over touch-tone dialling and reproduction of pre-recorded voice messages. In cases where human-machine dialogue includes selections among a small number of options and where it is possible to predict all possible answers or their combinations, there is no real need to include speech technologies. However, there are many more applications where the number of options is much greater and it is not practical or not possible to pre-record all the answers, and this is the area where speech technologies are of great use. For instance, if a speech technology application offers the user to listen to his/her e-mail messages, it is clear that the conversion of text into speech has to be carried out automatically, on-line. In the same way, an IVR application offering information related to a railway time-table would be much more practical if it allowed the user to ask for information using speech. Such applications are already quite common in countries where widely spoken languages are in use, but there are many countries where speech technologies have not yet been developed or are not yet in wide use.

II. Aids for Persons with Disabilities Based on Speech Technologies

Many people have impaired sight or hearing, a speech disability or some other psycho-physical limitation. Although it is not clear which degree of disability should qualify a person as disabled, it is estimated that nearly 10% of the whole population are persons with disabilities (PWD). Developed countries take good care of the people with disabilities. The UN charter on equal possibilities anticipated this fact 15 years ago, making it the subject of debates and a series of meetings of an ad-hoc UN committee aimed at making a draft of the "All-inclusive and integrative convention about promotion and protection of rights and dignity of persons with disabilities" [11]. EU, in its turn, initiated a number of programmes aimed at the improvement of the quality of life and the position in society of persons with disabilities. Although PWD are often limited by their incapability to use new technologies, these new technologies can be used to improve the aids for them and thus significantly improve their quality of life.

One of the recent EU initiatives is also e-inclusion [12], aimed at prevent a risk from being "digitally excluded", i.e. to make sure persons with disabilities are not neglected or overlooked because of their disability, lack of literature in an appropriate form or access to the Internet. At the same time, e-inclusion opens new possibilities for integration of marginalised groups and underdeveloped regions into the society. Information society provides all its citizens with more equal access to sources of information and opens new possibilities for employment. With the help of ICT, traditional barriers related to mobility and geographical dislocation are being broken. The "i2010 Communication" programme [13], aimed at an inclusive information society which should afford high quality public services and promote the quality of life, brings new challenges to the e-inclusion initiative. Finally, the year 2007 has been proclaimed as "the European year of equal possibilities for all", as an expression of the attempt to promote the rights and opportunities for persons with disabilities and to prevent their discrimination [14].

In accordance to European norms and standards in Serbia, the "Equal Rights for Persons with Disabilities" law has been adopted in 2006. The law proclaims equal rights for persons with disabilities, regarding education, information access as well as communication and employment. Nevertheless, many aspects of those rights will fail to be applied, unless modern technologies allow these persons to overcome their disabilities. Only then will they actualise their lawful rights to a more significant degree.

In the remaining part of this chapter it will be shown how speech technologies can help the visually impaired [15] as well as speech impaired, hearing impaired and physically impaired persons. Speech technologies can be applied as aids for persons with many types of disabilities. To the visually impaired, a machine can read books, newspaper articles from the Internet, e-mail or SMS messages. It can read aloud or talk instead of the speech impaired. It can receive voice commands from the physically impaired in order to control appliances such as telephone or any other household appliance connected to the system. It can convert input speech into text and thus serve as an aid to the hearing impaired.

A. The Speech Software for the Visually Impaired in WBC – anReader

Since the AlfaNum TTS [5] was presented for the first time, its popularity has been growing within the population of the visually impaired. Computers previously had to be used with TTS designed for foreign languages.

The AlfaNum R&D team subsequently implemented appropriate interfaces (SAPI 4 and SAPI 5) which made possible the use of AlfaNum TTS with any SAPI compatible screen reader. This speech software was named anReader. Highly intelligible and reasonably natural-sounding speech motivated blind people to use computers as speech machines, helping them to communicate and offering them access to information and literature, as shown in Fig. 2.1. In three years, the number of visually impaired computer users grew up to several hundreds within the population of about 13.200 visually impaired in Serbia. Adaptation for Croatian and Macedonian, doubled the population of anReader users in WBC.

The popularity of AnReader gave rise to training projects for our visually impaired compatriots, especially within a chain project named "Vizija", initiated by the Faculty of technical sciences. Several training centres for the education of the visually impaired were created, where experienced visually impaired computer users work as instructors.

Fig. 2.1. The blind can use computers as speech machines

As the anReader was declared a "valuable resource for the visually impaired", the Serbian Ministry of work and social politics provided many of the visually impaired computer users with a personal copy. At this moment, several thousands of the visually impaired in Serbia and other WBC have the opportunity for significant improvement of the quality of their lives.

B. Audio Library for the Visually Impaired

As the population of the visually computer users remained smaller than it could have been, within the AlfaNum project a new resource for the visually impaired was developed. The audio library for the visually impaired (ABSS v1.0) was developed for the pupils of the School for the visually impaired "Veljko Ramadanović" in Zemun, the largest education centre for this part of the Serbian population [16]. The lack of literature for the blind was a significant problem for the education process in the school. Until the end of 2005, education was mostly based on books in Braille as well as low quality speech synthesisers or audio-books. All of these were costly and took a lot of space. The audio library was developed with the aim of overcoming all of the mentioned problems.

The audio library is a client-server system containing a large amount of textual data that can be accessed by a number of users over the local network or, since recently, over the Internet. The server enables centralised importing of text and additional data, as well as sorting and searches. The client application is adjusted to the needs of the visually impaired and does not depend on any screen reader. Speech synthesis, based on the anReader, is used

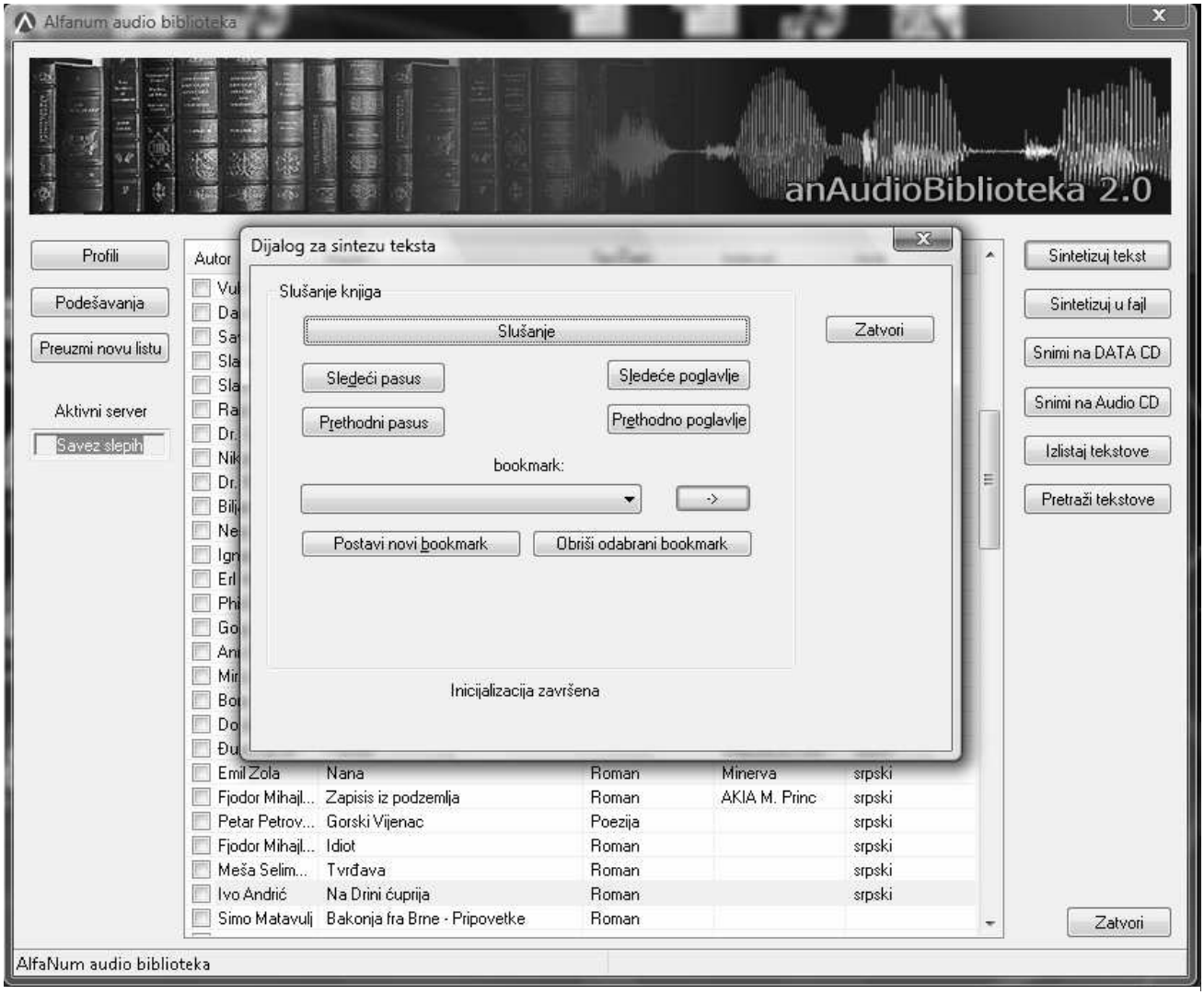

Fig. 2.2. The audio library for the visually impaired – a client application screenshot

for reading text as well as application commands. System offers the possibility of simultaneous access to the same book (all computers in the school being connected to the ABSS server). There is also a possibility of conversion of the book content to the audio format, as well as burning synthesised speech onto CDs or DVDs. In this way it can be used later by means of an ordinary CD/DVD player. A screenshot of the client application is shown in Fig. 2.2.

C. Voice Portal for the Visually Impaired

Newspapers in black print play a very significant role in our lives as a source of information and knowledge. However, black print is of little use for people with serious visual impairments without the help from somebody else. Having realised this problem, the experts from the Faculty of Technical Sciences and the AlfaNum company initiated a project named "Kontakt", and finished it by the first half of 2006. "Kontakt" is an Internet site which can be reached not only via web, but via a conventional telephone line as well (**0700/200 500**), making it accessible to the visually impaired people, regardless of whether they possess a PC or not. Internet access is based on a classic approach, through a web-browser, while phone access is based on recognition of spoken commands by the system. In both cases, site contents are read out using a speech synthesiser. Currently available content includes articles from web pages of several newspapers. Additional content will include more news sources, topics related to rights of persons with disabilities, entertainment, etc. A screenshot of the application is given in Fig. 2.3. The site can be found at http://www.alfanum.ftn.ns.ac.yu /kontakt. The web page is completely accessible by the visually impaired, since it is text-only, without pictures or banners, and the navigation is duly simplified. The primary design requirements were functionality and simplicity of use, rather than visual appeal, making it unconventional in comparison to a typical web site.

It can be concluded that now we have the opportunity to keep up with more developed countries and help the visually impaired in our community to improve their quality of life and be more actively involved in social life. The fact that the first steps were actually made in the biggest educational institution for the visually impaired people in Serbia, as well as in their unions, is of great importance, indicating at the same time a great interest for these innovative aids. Beside the audio library and the voice portal, the development of an SMS and e-mail reading system is under way. The development of these

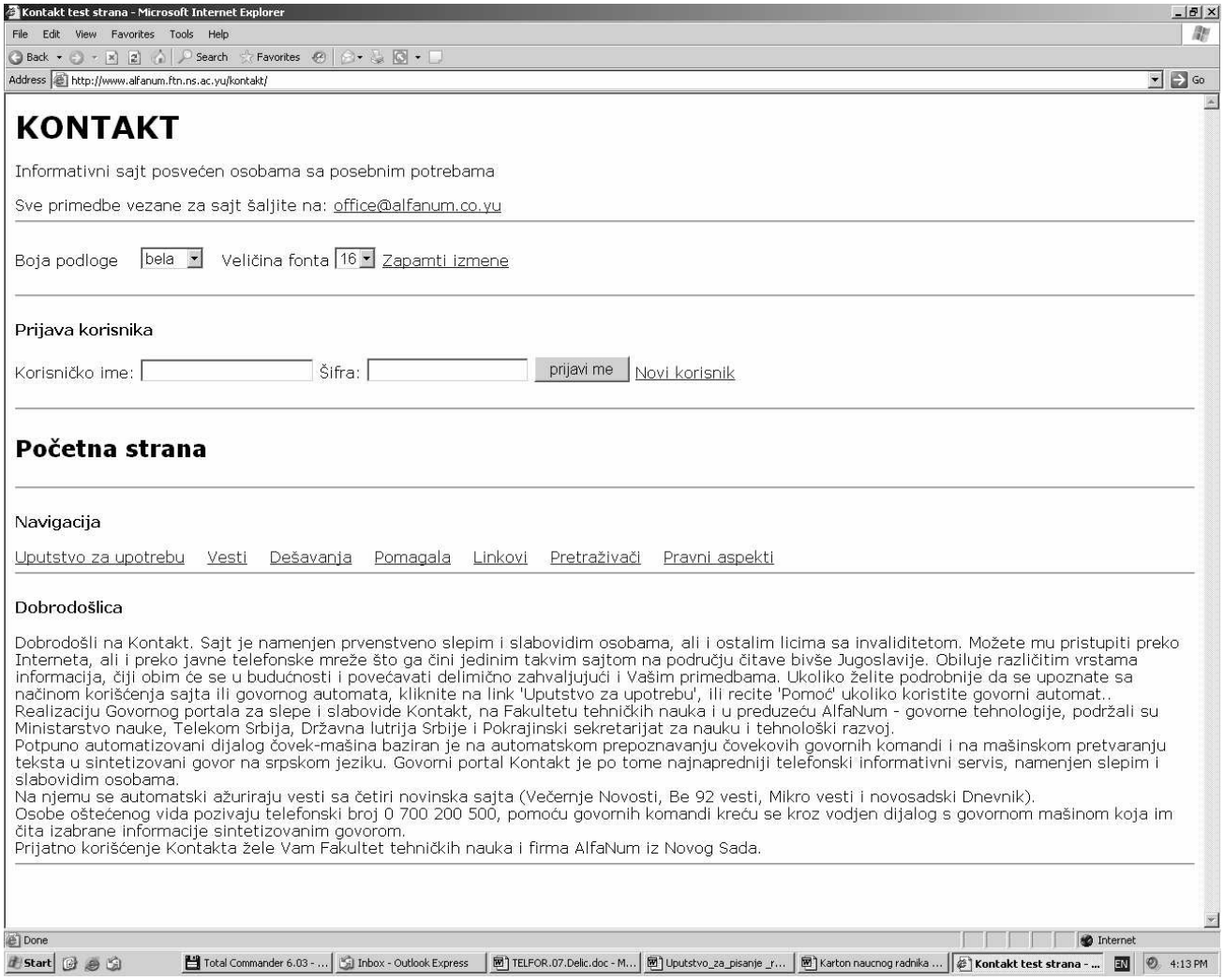

Fig. 2.3. The voice portal for the visually impaired "Kontakt" – a screenshot

important resources for the visually impaired should set an example for the development of similar resources for people with other disabilities, in areas where Serbian as well as other Slavic languages are spoken.

III. IVR Services Based on ASR and TTS in WBC

ASR and TTS are applied in telephony services within IVR systems. The basis of all IVR systems developed within the AlfaNum project is the simultaneous functionality of ASR and TTS servers and their communication with a required number of IVR processes (one per telephone line) via IP protocol, as shown in Fig. 3.1.

A unique database represents an information source from which data is presented to the user by TTS in the form of synthesised speech, based on user requests that the system acquires via ASR. The ASR and TTS servers can reside on remote computers (dedicated if required) and can communicate with a number of different IVR applications.

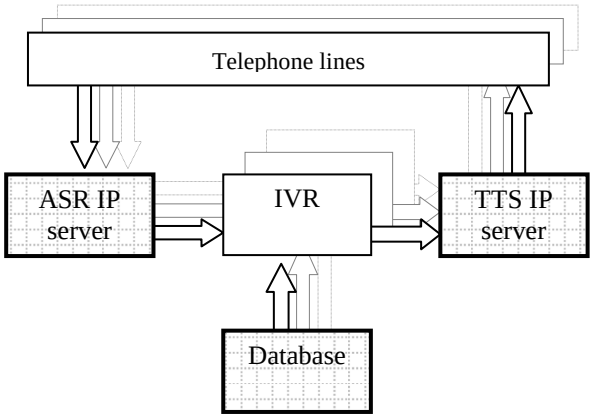

Fig. 3.1. IVR system based on ASR&TTS

A. IVR in Voice Portals

The first voice portals in South Slavic languages are the aforementioned portal for the visually impaired in Serbia ("Kontakt") and a similar in Croatia ("Televečernjak"). A fully automated human-machine dialogue is ASR and TTS based, as shown in Fig. 2.3. Daily information retrieval from various web sources is fully automated as well.

The experience acquired in the process of development and exploitation of these voice portals will contribute to the development of similar portals for other areas where South Slavic languages are spoken.

B. IVR in a Stock Market Application

This IVR system provides stockholders with information about their accounts at the Central Depositary Agency in Montenegro. Most of the application is standard, except for the recognition of isolated phonemes, since emission names are actually character strings such as URVF. Isolated phoneme recognition is unreliable as such; however, the number of combinations is drastically reduced due to the limited number of stock emissions that the clients actually possess. A dynamic design of grammars for a client's stock pool, without the need to restart the recogniser, was an additional requirement for the IVR system.

C. IVR in Entertainment Oriented Applications

"Sastanak" is an innovative and highly efficient phone dating IVR service. Human-machine communication enables users to execute criteria based searches through the database of registered users and to record voice messages intended for other users.

Selection of system functions (registration, search, playing and recording messages, ...) is carried out via ASR, while TTS is used for announcement of results of user actions (search results, data related to user profiles...), as well as outcomes of actual operations. Speech recognition is carried out by two independent ASR IP servers, which is sufficient for 30 telephone lines, while text-to-speech synthesis is carried out by two independent TTS IP servers.

IV. New ASR&TTS Applications in WBC

A. Multilingual Intelligent Telephone E-mail Access

This system enables its users to access their e-mail messages by phone, and is being developed by the AlfaNum team as well as the Alpineon from Slovenia [17].

Each day we spend more and more time reading e-mail messages. Most people also spend a lot of time travelling to work and from one business meeting to another, listening to music or radio news on their way. However, it is now possible to listen to e-mail messages via mobile phone, and thus save the time that would be spent at the beginning of the working hours. This has been made possible owing to recent improvements of the quality of text-to-speech conversion (TTS). An efficient phone e-mail access system should be able to handle a variety of e-mail messages in an intelligent way. An appropriate spam filter is also needed, as well as a simple navigation system (touch-tone and/or ASR based). The iTEMA system will offer personalization of the service regarding e-mail access as well as IVR communication, authentication and use of simple pre-defined replies.

Multilingual nature of the iTEMA system is characterised by language recognition at the sentence level and activation of an appropriate TTS engine. This is necessary owing to the language dependency of TTS, requiring that a TTS system for the appropriate language be used for synthesis. The iTEMA system will support reading e-mail messages in several widely spoken languages such as English, German, Italian (for which off-the-shelf speech synthesisers are available), as well as most South Slavic languages such as Slovene, Serbian, Croatian and Macedonian. Reading e-mail messages in some of these languages is a special challenge due to the fact that for writing messages in the same language one can use different scripts (Cyrillic or Latin, as well as Latin without diacritics).

The architecture of the iTEMA system contains an interface towards a number of SAPI compatible TTS engines. In the middle there is a dialogue manager connected to both telephone and Web interface. Personal settings for each user, such as mobile phone number, PIN, e-mail access parameters, are stored in a database.

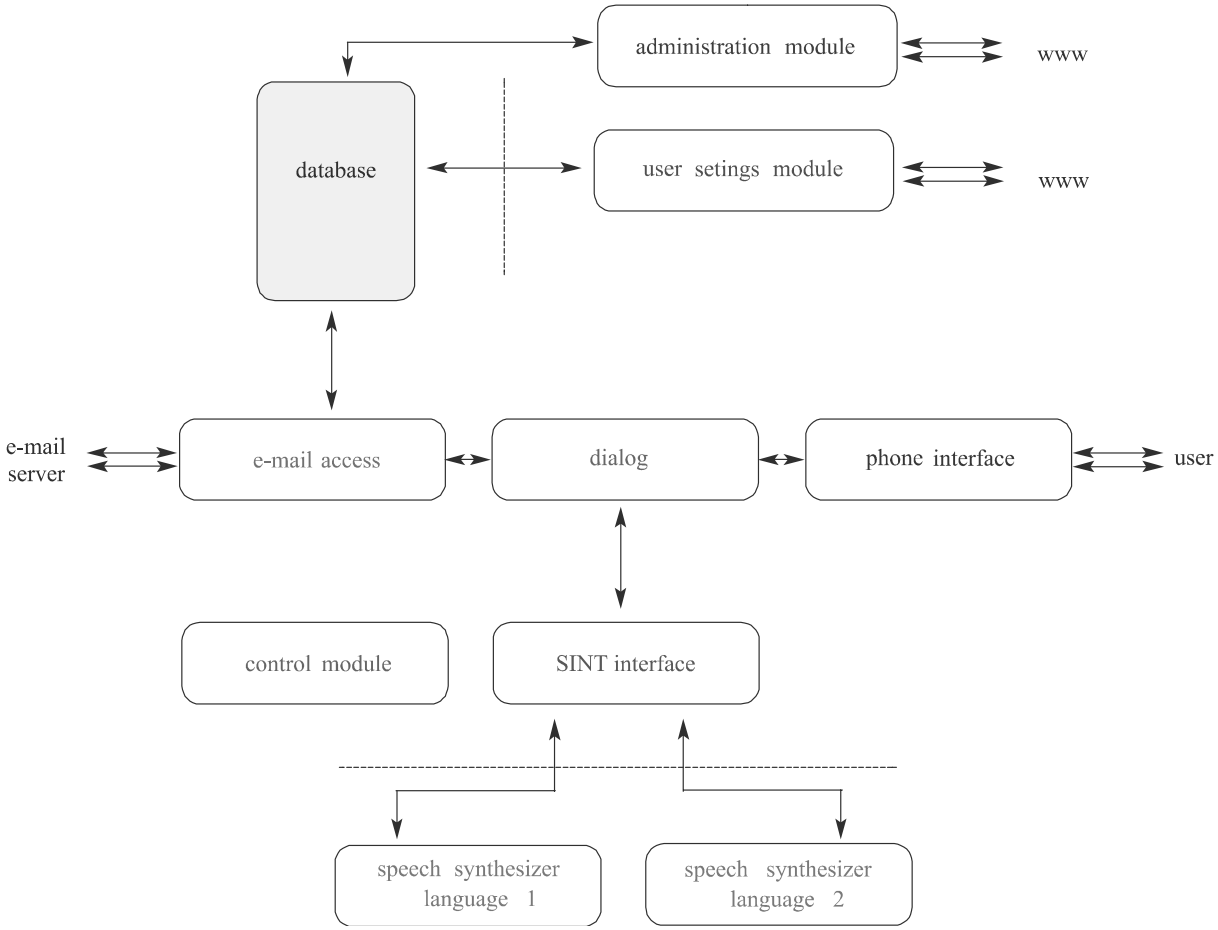

Fig. 4.1. System architecture of the iTEMA e-mail reader

A user dials the number of the iTEMA user service and a human-machine dialogue is initiated. Authentication is performed based on ANI and PIN, and followed by a personalised dialogue enabling simple and intuitive navigation through a menu system. Through this dialogue users can select messages they want to listen to, delete, or reply to using one of the pre-defined templates and recorded speech answer as an attachment.

Beside drivers and business people, iTEMA also provides e-mail service to those who have difficulties when using a computer but use a telephone as a matter of routine (the visually impaired, many of the elderly etc.). The iTEMA project thus represents material support to the e-inclusion programme of the EU.

Further improvement of the iTEMA service requires continuous improvement of the quality of synthesised speech, introduction of speech synthesis in a larger number of languages as well as successful language identification between them. It is also necessary to profile models of provision of the iTEMA service through cooperation with telecom operators, mobile telephony providers and ISPs.

Since speech communication is mostly one-way (from the TTS engine to the caller), potential delays in transmission of digitalised speech are not critical. For that reason, possibilities such as VoIP communication as well as GPRS transmission of synthesised speech can also be taken into consideration.

B. Network Based SMS-to-Speech Conversion

SMS messages converted to speech and read out aloud are more practical for users in case their hands and/or eyes are busy (e.g. while driving), if they do not see (the visually disabled) or check their SMS messages irregularly (the elderly). This application enables that an SMS be sent to a classic (fixed) telephone with no display unit, dials the number indicated at the beginning of the message and pronounces the message by means of synthesised speech. Such a service already exists in several countries. It has been made possible for Serbian in spite of the fact that letters with diacritics in the Latin version of Serbian alphabet (*š*, *č*, *ć*, *đ*, *ž* and *dž*) are usually coded in unconventional ways in SMS messages.

C. Advertising Monitor

Reliable automation of the process of monitoring radio and TV programmes can be achieved by means of ASR. Automatic recognition of audio recordings can be carried out very accurately because there is virtually no time variability, and the acoustic variability is minimal.

The Advertising Monitor [18] is a system for audio surveillance of commercial TV and radio programmes. It tracks several channels of TV and radio stations, records them in real-time and enables automatic tracking of particular commercials, jingles, music acts etc. It can be used by media monitoring enterprises or as a service offered to individual advertisers.

The Advertising Monitor consists of several FM and TV tuners as sources of audio signals (video excluded) for a sound card which can record 10 channels simultaneously. After on-line MSGSM compression, the recordings are archived on local hard drives. Depending on the number of channels and disk sizes, up to several months of audio material can be archived. Hence the Advertising Monitor is able to carry out retroactive tracking, since complete recordings of a number of radio/TV stations are stored for a period of several months. The Advertising Monitor records selected TV/radio signals according to the never-ending tape concept. Once the storage device is nearly full, the oldest recordings are automatically erased.

A search for a set of commercials defined by a sound file and channel settings can be executed through the archived recordings. The search is carried out by independent ASR processes activated on any of an arbitrary number of machines in the system. The results of the search are written into a shared database. This database is later used for creation of daily or monthly client reports, including an optional CD with a complete recording.

D. Word Spotting System

Manual search for keywords in lengthy audio recordings can be a very troublesome and time-consuming task. For that reason, automation of such process is very important.

Automatic recognition of predefined keywords and phrases in an arbitrary audio context (word spotting) cannot be done flawlessly. Due to a variety of contexts in which a keyword can appear, errors such as a false recognition or omission can occur. By raising the recognition reliability threshold for ASR, the number of false alarms will be reduced, but the probability of a keyword being missed will increase as a consequence. For each application an acceptable compromise has to be made.

Owing to automatic tracking of audio information in electronic media, firstly in TV broadcasts, studies of media influence on the public opinion are also possible. Long-term recording of all important radio and TV channels enables a retroactive audio tracking, which can be required by law in some cases as well.

E. Automatic Telephone Inquiry

The system for automatic telephone inquiry (ATI) can be used for automatic market analyses and opinion polls, by telephone, through an automated human-machine dialogue. The entire process of the inquiry is automated: inquiry design, asking questions and noting answers, data processing as well as report design [19].

A classic telephone inquiry is time-consuming and requires a lot of concentration from the inquirer, resulting in a variable quality in presenting the questions and reliability in noting the answers. ATI saves time and assets drastically, which means that small businesses as well could afford market analyses and public opinion polls, as well as those who wish to keep track of quick changes in the market or public opinion.

Beside the ATI application itself, applications for inquiry preparation, review and verification of answers and statistical analysis as well as graphical representation of

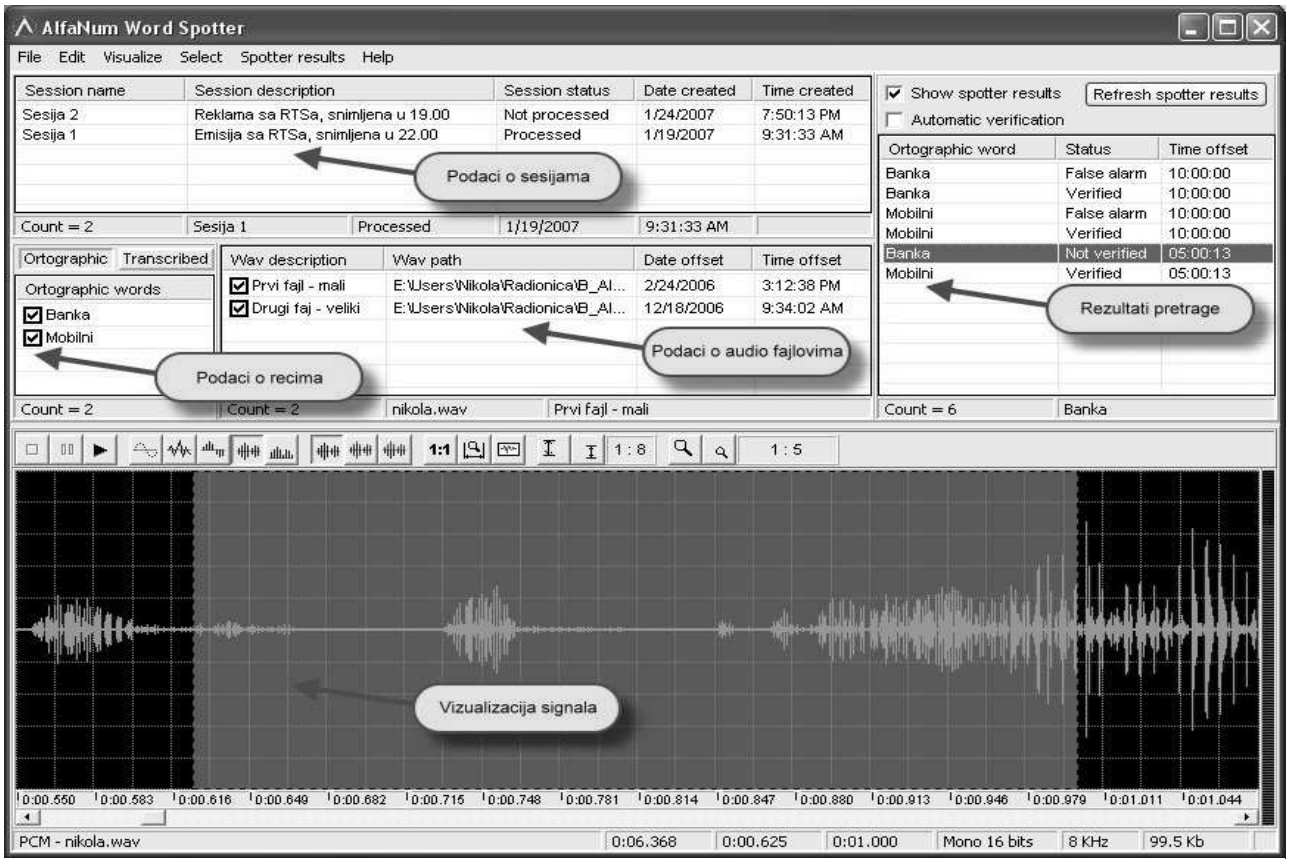

Fig. 4.2. GUI of the Word Spotting System

results are developed. ATI users can carry out a statistical analysis at any time, making it easier to identify issues which require a modification of the inquiry, presented questions or multiple choice answers.

V. Conclusion

The Serbian language belongs to a relatively small circle of languages in the world for which speech technologies reached a level of quality necessary for their initial wider application.

We are being witnesses of a constant increase in ASR accuracy and robustness, and in intelligibility and naturalness of TTS. In areas where Serbian, Croatian, Bosnian and Macedonian are spoken, the first applications were aimed at people with disabilities, and subsequently expanded to the CTI market, mainly as IVR systems. In the meantime, innovative applications such as the Advertising Monitor and Word Spotter were created. Horizons are expanding for different applications of ASR and TTS, like Serbian language learning for foreigners or even automatic speech translation from Serbian to other languages with ASR and TTS support, directly or (more likely) using English as an intermediary.

ASR and TTS are complex technologies which will most likely never function flawlessly, but they always function with constant quality and often more efficiently than human beings – at present, a single PC can talk to tens of clients simultaneously – to understand their voice commands and to provide them with information via speech.

Current research is expanding in the direction of speaker recognition, emotional state detection, and machine translation of speech. It is extraordinarily important that each of us strives for realisation of these applications in one's own language as well.

Acknowledgment

This work was supported in part by the Ministry of Science of Serbia within the Project "Development of speech technologies in Serbian and their application in 'Telekom Srbija'" (TR-6144A).

References

[1] V. Delić, D. Pekar, R. Obradović, M. Sečujski, "Speech Signal Processing in ASR&TTS Algorithms," *Facta Universitatis,* Vol. 16, No. 3, Dec. 2003, pp. 355-364 [TELFOR, 2002, p. 454].

[2] V. Delić, M. Sečujski, D. Pekar, N. Jakovljević, D. Mišković, "A Review of AlfaNum Speech Technologies for Serbian, Croatian and Macedonian," *in Proc. Int. Language Technologies Conf.,* IS-LTC, Ljubljana, Slovenia, 2006, pp. 257-260

[3] V. Delić, M. Sečujski, Ž. Tekić, "A Contribution to Human-Machine Communication in Serbian, Croatian and Macedonian Language," *in Proc. 17th Int. DAAAM'06 Symposium, Intelligent Manufacturing & Automation,* Vienna, Austria, 2006, pp. 101-102

[4] S. Sovilj-Nikić, M. Sečujski, V. Delić, I. Sovilj-Nikić, "Analysis of Different Factors Influencing Vowel Duration in Serbian Language," *in Proc. Inte. Symposium on Social Communication,* Santiago de Cuba, Cuba, 2007, pp. 385-389

[5] M. Sečujski, R. Obradović, D. Pekar, Lj. Jovanov, V. Delić, "AlfaNum System for Speech Synthesis in Serbian Language," *in Proc. Text, Speech and Dialogue,* Brno, Czech Republic, 2002, pp. 237-244

[6] M. Sečujski, "Prozodijski elementi u sintezi govora na osnovu teksta na srpskom jeziku," MSc thesis, University of Novi Sad, Serbia, 2002.

[7] M. Sečujski, "Obtaining Prosodic Information from Text in Serbian Language," *in Proc. IEEE EUROCON,* Belgrade, Serbia, 2005, pp. 1654-1657

[8] R. Obradović, D. Pekar, S. Krčo, V. Delić, V. Šenk, "A Robust Speaker-Independent CPU-Based ASR System," *in Proc. EUROSPEECH,* Budapest, Hungary, 1999, pp. 2881-2884

[9] M. Gilbert, J. G. Wilpon, B. Stern, G. Di Fabbrizio, "Intelligent Virtual Agents for Contact Centre Automation," *IEEE Signal Processing Magazine,* Vol. Sept. 2005.

[10] D. Pekar, R. Obradović, V. Delić, "AlfaNumCASR – a System for Continuous Speech Recognition," *in Proc. DOGS,* Bečej, Serbia, 2002, pp. 49-56

[11] http://www.un.org/esa/socdev/enable/rights/adhoccom.htm

[12] http://europa.eu.int/information_society/soccul/eincl/index_en.htm

[13] http://europa.eu.int/information_society/industry/comms/index_en.htm

[14] http://ec.europa.eu/employment_social/equality2007/index_en.htm

[15] V. Delić, N. Vujnović, M. Sečujski, "Speech-Enabled Computers as a Tool for Serbian Speaking Blind Persons," *in Proc. IEEE EUROCON,* Belgrade, Serbia, 2005, pp. 1662-1665

[16] D. Mišković, N. Vujnović, M. Sečujski, V. Delić, "Audio biblioteka za slepe i slabovide osobe kao vid primene TTS tehnologije," *in Proc. ETRAN,* Budva, Montenegro, 2005, Tom II, pp. 400-402

[17] J. Žganec Gros, V. Delić, D. Pekar, M. Sečujski, A. Mihelič, "The iTEMA E-Mail Reader," *in Proc. Int. Language Technologies Conf.* IS-LTC, Ljubljana, Slovenia, 2006, pp. 230-233

[18] D. Pekar, G. Kočiš, R. Vuković, S. Molerov, "AlfaNum Advertising Monitor," *in Proc. TELFOR,* Belgrade, Serbia, 2006.

[19] V. Delić, N. Vujnović, B. Kostić, "Mogućnosti automatizacije telefonskih anketa pomoću govornih tehnologija," *in Proc. DOGS,* Vršac, Serbia, 2006, pp. 63-66

[20] V. Delić, "A Review of R&D of Speech Technologies in Serbian and their Applications in Western Balcan Countries", *Keynote lecture at XII international conference "Speech and Computer" (SPECOM),* Moskva, 15-18. Oct. 2007. pp. 64-83

[21] V. Delić, D. Pekar, R. Obradović, N. Jakovljević, D. Mišković, "A Review of AlfaNum Continuous Automatic Speech Recognition System", *in Proc. XII international conference "Speech and Computer" (SPECOM),* Moskva, 15-18. Oct. 2007, pp. 702-707

[22] Milan Sečujski, Vlado Delić, Darko Pekar, Radovan Obradović, Dragan Knežević: "An Overview of the AlfaNum Text-to-Speech Synthesis System", *in Proc. XII international conference "Speech and Computer" (SPECOM),* Moskva, 15-18. Oct. 2007, pp. A3-7

[23] http://www.alfanum.ftn.ns.ac.yu

[24] http://www.alfanum.co.yu

ASR & TTS APPLICATIONS IN WESTERN BALKAN COUNTRIES

Vlado Delić