Building of the annotated speech utterances database from the Slovak spoken dialog system interactions

Matúš Pleva, Jozef Juhár, Anton Čižmár

Department of Electronics and Multimedia Communications,
Technical University of Košice, Park Komenského 13, 041 20 Košice,
Slovak Republic
Email: Matus.Pleva@tuke.sk, Jozef.Juhar@tuke.sk, Anton.Cizmar@tuke.sk

Abstract

The development of the spoken dialog system in Slovak is not finished after releasing the functional version. As a human needs to adapt to new words, pronunciations and grammar during the life also a dialog based server needs to adapt using the realized interactions logging. The analysis and annotation of the recorded speech utterances from the spoken dialog system interactions could provide the improvement of the acoustic models, grammars and dictionaries according to real needs of the system users. This article describes the building of the database from recorded dialog interactions with real users. The logged audio data with ASR (automatic speech recognition) recognized words are the sources for developed transcription tool. The manually corrected transcriptions and the audio data are then converted to desired database format, suitable for the ASR training.

1. Introduction

The goal of the approach is to build a database of the recordings compatible with databases that are commonly used for ASR training – building dictionaries, grammars, acoustic models, etc. The desired format for the chosen spoken dialog server in Slovak language [1] is the SpeechDat-E database format [2].

Firstly the recorded utterances and the logs from the recognized words are periodically converted and transferred to the annotators' server. Then the logs are manually corrected and then converted to the SpeechDat-E format and stored in the ASR training database server.

2. Collecting the utterances

During the years of providing information about weather and train connections the spoken dialog system recorded gigabytes of speech utterances of real human-computer interactions in Slovak language [3].

This data are important for improving the robustness of the human-computer dialogs. The way to improve the dialog is firstly to annotate the human part of the dialog and then to analyze the results.

The ASR server of the communicator is storing all audio data used for automatic speech recognition from the user utterances. This audio data are stored together with the recognized text (stored from dialog manager server) after the hang-off.

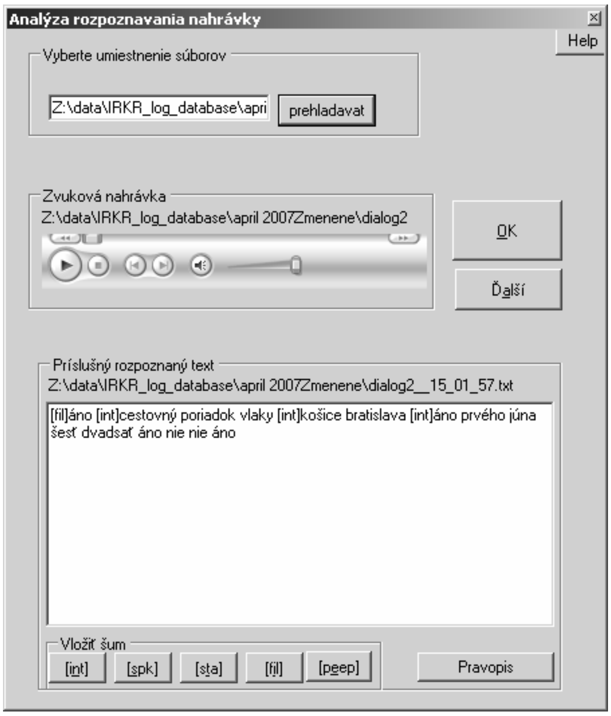
Using Perl scripting language the logged data are converted to a wav files and a text files with user part of the interactions (the original log file contains the whole dialog).

3. Correcting the annotations of the utterances

The preliminary annotation is taken from the recognized texts of the ASR server. This data speeds up the annotation process. In the recordings, there are sometimes also recordings of the hang-off tone from the phone line. This tone is annotated with time stamp, which could bring the option to cut-off the tone part of the recording during the conversion to the SpeechDat-E database format.

The transcription is done without time stamps, because the user part of the interaction is usually not very long, and the ASR server recognizes the words from the grammar that are usually not longer than 3 words (fixed grammar recognition). The longer utterances are now not included in the training database. In the future, we plan to add a timestamp feature also to the transcribed text.

The transcription tools loads the data from the directory and after correcting the transcription they are stored and after automatic spell checking the next files are loaded immediately as we can see on the Figure 1.



Vyberte umiestnenie súborov: Z:\data\IRKR_log_database\apri

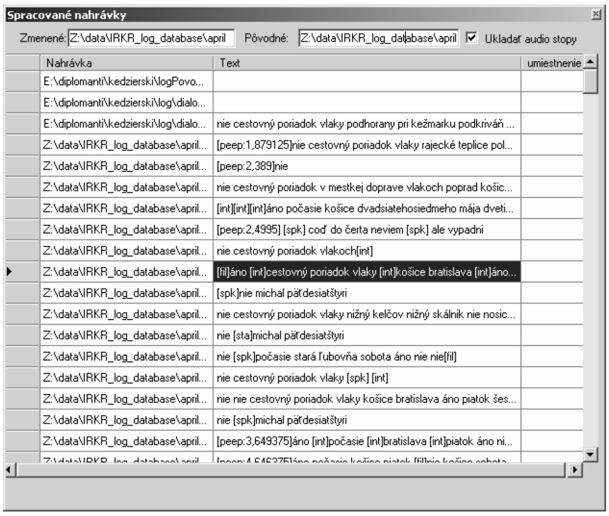
Zvuková nahrávka: Z:\data\IRKR_log_database\april 2007Zmenene\dialog2

Príslušný rozpoznaný text: Z:\data\IRKR_log_database\april 2007Zmenene\dialog2__15_01_57.txt

[fil]áno [int]cestovný poriadok vlaky [int]košice bratislava [int]áno prvého júna šesť dvadsať áno nie nie áno

Figure 1. Transcription tool main window with automatic spell checking feature

The tool was built using Visual Basic and the spell checker is loaded from MS Office libraries. The second window of the tool provides information about the files in the transcribed directory with relevant transcriptions (see Figure 2). There we can choose if the audio data are also stored with the corrected transcriptions in the new directory and we can quickly choose the next file we want to correct.



Zmenené: Z:\data\IRKR_log_database\april Pôvodné: Z:\data\IRKR_log_database\april ☑ Ukladať audio stopy

Nahrávka	Text	umiestnenie
E:\diplomanti\kedzierski\logPovo...		
E:\diplomanti\kedzierski\log\dialo...		
E:\diplomanti\kedzierski\log\dialo...	nie cestovný poriadok vlaky podhorany pri kežmarku podkriváň ...	
Z:\data\IRKR_log_database\april...	[peep:1,879125]nie cestovný poriadok vlaky rajecké teplice pol...	
Z:\data\IRKR_log_database\april...	[peep:2,389]nie	
Z:\data\IRKR_log_database\april...	nie cestovný poriadok v mestkej doprave vlakoch poprad košic...	
Z:\data\IRKR_log_database\april...	[int][int][int]áno počasie košice dvadsiatehosiedmeho mája dveti...	
Z:\data\IRKR_log_database\april...	[peep:2,4995] [spk] coď do čerta neviem [spk] ale vypadni	
Z:\data\IRKR_log_database\april...	nie cestovný poriadok vlakoch[int]	
Z:\data\IRKR_log_database\april...	[fil]áno [int]cestovný poriadok vlaky [int]košice bratislava [int]áno...	
Z:\data\IRKR_log_database\april...	[spk]nie michal päťdesiatštyri	
Z:\data\IRKR_log_database\april...	nie cestovný poriadok vlaky nižný kelčov nižný skálnik nie nosic...	
Z:\data\IRKR_log_database\april...	nie [sta]michal päťdesiatštyri	
Z:\data\IRKR_log_database\april...	nie [spk]počasie stará ľubovňa sobota áno nie nie[fil]	
Z:\data\IRKR_log_database\april...	nie cestovný poriadok vlaky [spk] [int]	
Z:\data\IRKR_log_database\april...	nie nie cestovný poriadok vlaky košice bratislava áno piatok šes...	
Z:\data\IRKR_log_database\april...	nie [spk]michal päťdesiatštyri	
Z:\data\IRKR_log_database\april...	[peep:3,649375]áno [int]počasie [int]bratislava [int]piatok áno ni...	

Figure 2. Transcription tool second window with file list and directories

4. Converting to desired database format

All the corrected transcriptions are saved in the directory of corrected logs. After finishing the corrections, the text files needs to be converted to the SpeechDat-E SKO format (SAM 6.0 [4]). The audio files are converted using sox open source tool [5] to raw header-less 16 bit PCM format.

The SAM description file contains all available information about the speaker (gender), the recording time and date (from log file name), transcribed text, raw file description and network source (telecommunication network from which the interaction was recorded and a possible codec used as it is depicted in the Table 1.). In the future also a caller phone number will be stored using CLIP (Caller Line Identity Presentation) information from the providers.

“Table 1. Dialog source table”

Log file stamp	Dialog source (codec)
dialog1	PSTN (PCM-Alaw)
dialog2-4	VoIP (H323, PCMA)
dialog5	VoIP (Skype)
dialog6	GSM (T-mobile)
dialog7	GSM (Orange)
dialog8tcom	GSM (t-mobile)
dialog9sip1	VoIP (sip, PCM-ulaw, G722, G729)

The files are stored in the directory structure also followed by the recommendations of the SpeechDat-E project. This directory structure could use the training scripts as a source for new or retrained acoustic models training.

5. Future plans

This new database covers application specific data, collected during the human-computer interaction with computer initiative dialog using fixed grammar and dictionary. It means that the data could be effectively used for increasing the robustness of the spoken dialog server in Slovak language for these specific applications. For example: for increasing the recognition accuracy for most frequently used train or bus stations in the travel timetable application.

This new database of corrected speech utterances recorded from spoken dialog human-machine interaction will be used for:

1. training the acoustical models for automatic speech recognition in Slovak
2. modifying grammars according to user replies on the dialog questions (new possibilities to say yes/no, etc.)
3. modifying dictionaries and including new words or new word forms (slang, dialect, …)
4. the annotations could be used for training the language model for continuous speech recognition in Slovak language (in case that user produce an out of grammar speech utterances annotated during the correction procedure)
5. retraining existing acoustic or language models and compare the results in recognition accuracy

After collecting a significant volume of transcribed and corrected data, the database will be tested and then joined with the databases used for spoken dialog system development in Slovak language.

7. Conclusion

This corrected recordings database is useful for development of spoken dialog systems, also because it contains spontaneous reactions of the humans on the computer initiative dialogs.

This database collecting procedure, including Perl scripts, transcription tool and log collecting mechanism, is a guide to extending the domain-specific speech materials with less energy given to transcription process. The domain specific data (aprox. 3 hours) extracted from the database could increase the robustness of the spoken dialog system by giving the information to the dialog designers about the actual new words, new word or sentence forms and new dialog routines which could be included to the computer-human interaction.

8. Appendix and acknowledgments

The work presented in this paper was supported by the Ministry of education of Slovak Republic under research projects AV 4/0006/07, MVTS COST2102/07, VEGA 1/4054/07 and AV 4/2016/08 and Slovak Research and Development Agency under research project APVV-0369-07.

References

[1] Juhár, S. Ondáš, A. Čižmár, M. Rusko, G.Rozinaj, R. Jarina, “Development of Slovak GALAXY/VoiceXML Based Spoken Language Dialog System to Retrieve Information from the Internet”, *Proc. of ICSLP, Pittsburgh*, USA, September 17-21, 2006, pp. 485-488.

[2] M. Rusko, S. Daržagín, M. Trnka, “SpeechDat-E telephone speech database as an important source for basic acoustic-phonetic research in Slovak”, in *Proceedings of ICA 2004*, Kyoto, Japan, 2004, pp. II-1676-II-1682.

[3] http://irkr.tuke.sk/

[4] M. Tomlinson, R. Winski, W. Barry, “Label file format proposal”, from *Esprit project 1541 (SAM): Extension Phase - Final Report*, 1988, p.189-197

[5] http://sox.sourceforge.net/