Speech based Drug Information System for Aged and Visually Impaired Persons

Géza Németh, Gábor Olaszy, Mátyás Bartalis, Géza Kiss, Csaba Zainkó, and Péter Mihajlik

Department of Telecommunications and Media Informatics
Budapest University of Technology and Economics, Hungary
{nemeth, olaszy, bartalis, kgeza, zainko, mihajlik}@tmit.bme.hu

Abstract

Medicine Line (MLN) is an automatic telephone information system operating in Hungary since December 2006. It is primarily intended for visually handicapped persons and elderly people. In Hungary the National Institute of Pharmacy (NIP) coordinates the approval of new drugs and also their Patient Information Leaflets (PIL). Medicine Line reads this textual information chapter by chapter to the citizens. The number of different medicines used in Hungary is about 5000. New drugs come into practice regularly, and some are withdrawn after a certain time. The MLN system ensures 24 hour access to the information. The spoken dialogue input is processed by a specialized ASR module (the caller tells the name of the drug, the chapter title etc.). The output is given by a TTS synthesizer specialized to read drug names and medical Latin words correctly. The user can control the system by DTMF buttons too. In this article we will focus on features of speech based components.

Index Terms: medical TTS, ASR for drug names, combination of TTS and ASR, pronunciation of drugs, Patient Information Leaflet, medical texts, drug information

1. Introduction

Using speech technology in medical fields is a developing field. Radiologists can already dictate their analyses of X-ray pictures, and the spoken words will be transcribed into text. Physicians can dictate prescription orders into wireless hand-held devices and the ASR enters the spoken items into text which can be saved after confirmation into a central information system for later use [1]. Investigations show, that ASR – in spite of its present accuracy rate (80%-95%) – is being used more frequently in dictating medical reports [2].

One of the main problems of using speech technology in the medical field is to process correctly by software the special language elements (written or spoken) of this field, including the terminology (drug names, Latin words etc.) [3]. In this article the combination of ASR and TTS is presented for a drug information system. The system is to be used mainly by the public.

The National Institute of Pharmacy coordinates the approval of new drugs in Hungary. The number of different medicines used is about 5000. Pharmaceutical manufacturers submit a PIL about the drug the text of which is approved by NIP. This leaflet contains information about the main use of the product, what is important to know before the use, how to use, what side effects can occur, etc. New drugs come into practice regularly, and there are some which are withdrawn after a certain time, so the fluctuation of the products is rather high.

Two aims have been taken into consideration by design of this pharmaceutical information system. As to the first, to allow 24 hour access for visually handicapped persons and for elderly people to the text of the patient information leaflet of every drug. The system besides can help physicians and chemists in their work, too. The second important aim was to design and develop a user friendly, easy to use system for all. That is why the spoken dialogue solution was chosen. A user independent specialized ASR module accepts the voice of the caller who can define the medicine in question by telling its name. After recognition and confirmation of the drug name the information leaflet of the drug is read by a TTS synthesizer specialized to pronounce drug names, medical Latin words and special expressions correctly. The dialogue between the caller and the machine is provided by voice all the time.

2. System components

The Medicine Line system has five main components, which will be detailed below. ASR for drug names, TTS for reading the text of PILs, dialogue controller, database and the automatic updater (Figure 1).

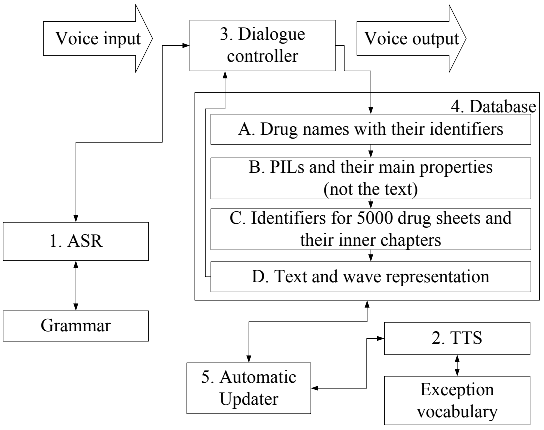
Voice input
3. Dialogue controller
Voice output
4. Database
A. Drug names with their identifiers
B. PILs and their main properties (not the text)
C. Identifiers for 5000 drug sheets and their inner chapters
D. Text and wave representation
1. ASR
Grammar
2. TTS
5. Automatic Updater
Exception vocabulary

Figure 1: *The main blocks of the Medicine Line system.*

2.1. ASR for drug names

The main goals of the ASR for drug names are: to recognize the spoken drug names at telephone quality with more than 95% accuracy; to ensure an easy form to define new drug names; to recognize also the user commands with high accuracy to control the dialogue. Specialized grapheme to phonemes rules are applied for automatic pronunciation model generation. The operator can manually control the phonemic transcripts. Phoneme acoustic models are taught on a SpeechDat-like Hungarian speech database, with approximately 20 hours speech [4][5]. An MFA-based front-end [6] is used to get MFCC-based acoustic features [7]. Speaker independent decision-tree state clustered cross-word triphone models with approximately 2000 HMM states were trained using ML estimations [8]. Three state left-to-right structure HMMs were applied with 10 mixture components per state.

Since users are forced to tell only the name of medicine no word spotting or continuous speech recognition is applied. Therefore no language model, or in other word, zero-gram model is used. The vocabulary is about 5K words. To speed up the recognition process the recognition network is optimized off-line at HMM state level using WFST algorithms [9]. One-pass decoding is performed by an enhanced version of the decoder described in [10] close to real-time on a 3GHz CPU up to 32 channels load.

2.2. TTS for pharmaceutical texts

The TTS module of the MLN system was developed from the Hungarian Profivox synthesizer [11]. The Profivox-Med reading software has two special features compared to Profivox. First, it is capable to pronounce thousands of drug names, and Latin words occurring in the PILs correctly. Second, it is prepared to synthesize (mainly from the prosodic point of view) the special style used by pharmaceutical manufacturers when writing the PILs. It must be emphasized, that the texts of PILs are not composed for TTS conversion, but for human reading. The style is difficult and there are also many spelling mistakes. Humans jump over the misspellings during reading and interpreting the text, but the TTS converts them letter by letter, as they are written, so the sound will be incorrect. In our opinion manufacturers should be convinced to take into account in the future the speech technology demands in this field. This is their interest, too.

2.2.1. Reading medical terminology by TTS

During the preparation of the Profivox-Med software for correct reading of medical terminology all texts of all PILs have been processed to pick out and collect the non-Hungarian words and text parts occurring in the PILs. The majority of the selected words were drug names and Latin words, but other items like foreign words, abbreviations, chemical expressions have been found also (N-hepa; hidroxi-propil-metil-cellulóz; 40 µg PGE; 800 mOsm/1; kallikrein inactivator unit; HMG-CoA; non-Hodgkin lymphoma etc.). The list has been examined and a rule based pronunciation sub-module was developed supported by a special exception vocabulary to reach the correct TTS conversion of these items. The rule based part touched mainly the drug names and the Latin words (Table 1). In this part pronunciation rules have been defined for certain letter combinations. The number of these rules is 243.

Table 1. *Example of pronunciation rules for drugs for TTS conversion.*

drug name	Pronunciation	word end rule
ACCOLATE	[akkolat]	te# = [t]
ACICLOSAN	[a ts iklozan]	san# = [z a n]
ACTILYSE	[aktiliz]	yse# = [i z]
ALKA-SELTZER	[alkasel ts er]	tzer# = [ts e r]

The exception vocabulary contained the transcription of the remaining elements by sound symbols (word by word). The use of this combined solution in the TTS system resulted in a correct reading of medical terminology embedded in the text. Enlarging the vocabulary is continuous, because in the case of new drugs new items can occur with wrong pronunciation. In this respect, this part of the system is open and needs continuous manual work from the operator.

2.2.2. Special style of the PILs

Patient Information Leaflets are originally written for humans, not for a machine. As every profession, the pharmatheutical industry has a special style for these texts independently of the manufacturer. First of all in many cases there are very long sentences, and their structure is complicated. A human reader understands the structure of the text, so the interpretation, the perception of the essential parts is not a problem. To read this by a machine needs the use of special pausing and intonation strategy in the TTS converter. Embedded medical expressions in the text make the understanding more difficult too.

The sentences also have in many cases text parts between brackets (references to drug names, patient indications etc.). The application of special prosody is needed to make perceptible by voice that these text parts are between brackets.

Many sentences contain long enumeration, when for example side effects are given in the case of which the use of the drug is not proposed. See the next sentence: *If you feel side effects, as for example squeamishness of stomach, sweating, shaking, weakness, giddiness, dryness in the mouth, sleepiness, sleeplessness, costiveness, diarrhea, less appetite, nervousness, excitement, headache, sexual troubles, please ask your doctor to modify the dosage.* To find in the text the enumeration parts is sometimes not easy because in many cases separators (comma) are not used, only the text continues in a new line. A pause module has been developed to handle such long enumerations. The special prosody module of the Profivox-Med TTS system handles all these problems.

During laboratory level evaluation of this module hundreds of PILs have been listened to and criticized by one expert, and a phonetician constructed new rules for the TTS module based on the remarks. By relistening the same texts less remarks has been given by the test person. For these remarks new rules have been constructed and so on. Finally this module got 34 new rules for improving the original pausing strategy and prosody rules have been extended by 12 new rules also. The use of this module made the synthesized pharmaeutical text more understandable in the majority of cases. It must be emphasized that the full solution of this problem cannot be solved until machines will not be capable of understanding the meaning of the text. For example in many cases a pause is needed before or after a reference between brackets, but in other ones this pause is disturbing.

2.2.3. Text preparation for TTS conversion and database storage

Two text representation forms are applied: chapters and sentence selection.

Chapters serve to divide the text of PILs into smaller parts. Why was it needed? The overall text of PILs is in most cases rather long. To read this long text to the patient as one unit is not the most optimal solution. Patients may want to hear only one part of the leaflet. The NIP of Hungary also earlier suggested dividing the text of the PIL into chapters. According to this, every PIL contains 5 chapters such as *What the drug is good for? Before using. How to use? Side effects. How to store?* The TTS conversion was organized also by these chapters. So the dialogue controller offers the chapter titles for the user, and after selection the system will read only the selected chapter. The actual selection by the user is made by voice, the chapter title is pronounced by the patient.

Sentence selection is also a special feature when preparing the texts of PILs for TTS conversion. As there may be many equal sentences in the texts of all PILs, it was found that it is worthwhile, to store each equal sentence only once in the database. To make this selection the sentences of all PILs were compared and a sentence store was constructed in which only no textually equal sentences have been placed. All these sentences have a unique identifier. This store is placed in part D of the database. Parallel with the sentences their synthesized waveforms are also stored with the same identifier.

2.3. Spoken dialogue controller

The dialogue controller is perhaps the most sensitive part of a spoken dialogue system. The dialogue between man and machine should be optimal i.e. not complicated but user friendly and clear to use. Therefore the deepness level of the menu structure is only three. The goal is to find the drug of interest as quickly as possible. The dialogue begins (after the basic welcome) with the speech of the caller who tells the name of the drug she/he wants to select from the database. Only the name of the medicine should be told, it is in most cases only one word i.g Aspirin (more words are not allowed eg. "Aspirin tablets"). The ASR module selects the drug, and as a feedback the TTS tells the name of the drug back to the user, who confirms the selection. As the second step, the user selects by voice the desired chapter of the PIL, by repeating the prespoken chapter name. The TTS begins to read the chapter. The Stop/continue speaking, repeat the sentence, jump functions make easier for the user to navigate inside the synthesized text of the chapter. These are controlled by buttons of the phone.

The dialogue controller also handles the case when more items are found in the database under the same drug name (Table 2).

Table 2. *Examples of having several items under the same drug name in the database. In case of Aspirin 7 items, in case of Algopyrin 3 versions are available.*

The spoken drug name	The items found in the database
ASPIRIN	ASPIRIN 100 pill ASPIRIN PLUS C effervescent tablet ASPIRIN DIREKT chewing pill ASPIRIN 500 pill ASPIRIN MIGRAIN effervescent tablet ASPIRIN PROTECT 100 mg intestine-soluble pill ASPIRIN PROTECT 300 mg intestine-soluble pill
ALGOPYRIN	ALGOPYRIN 1g/2ml injection ALGOPYRIN 500 mg pill ALGOPYRIN COMPLEX pill

In this case the dialogue controller asks the user to select one of them by voice, telling the order number of the item which is prespoken by the TTS. Finally a "context sensitive help" is also for the disposal of the user at every level of the menu to make the dialogue more user friendly.

2.4. Database

Four groups of data are stored in the database (Figure 1.). Part A contains the drug names with their identifiers; part B has the main marginal data of the PILs and the drugs; part C contains the identifiers for PILs, their chapters and the sentence identifiers; part D has the sentence store i.e. the selected sentences in text and wave formats.

Part A consists of data for the 5000 drugs, and the dialogue controller transmits the decision of the ASR here. The selection of the drugs is done here. In part B the PIL is selected, part C will define exactly (on identifier level) which PIL and which chapter of it will be put on the voice output of the system. Part D organizes the voice output stream on the basis of the identifiers of C. The synthetic voice will be put on the voice output by the dialogue controller. The user can ask for normal or faster speed in reading (e.g. visually impaired like faster speech speed).

2.5. Automatic updater

The automatic updater (Block 5) is a web based administration tool to help refreshing the data of drugs (update the database) regularly, twice in a month. This is needed because there are changes continuously in the list of drugs (new drugs are added to the list, some of them are to be deleted). This update is done automatically. The operator is always preparing a loader file, which contains the actual name of the drugs, and the file names of the PILs. During updating a comparison is done. The name of a new drug found in the loader file, will be put into the database and the TTS generates the pronunciation form of its name. If a new PIL is found, the updater compares the sentences of the PIL with the sentences of the sentence store. If a new sentence is found the TTS converts it into speech. The text of this new sentence and the waveform of it's synthesized version will be added to the sentence store (database part D). The PIL chapter identifiers and the basic information of the PIL will be added too (parts B and C).

The loader file contains several information elements about the new or changed drugs like:

-The brand name of the drug (e.g. ALGOPYRIN)
-The full name of the drug (e.g. ALGOPYRIN 500 mg pill)
-The main active substance of the drug (e.g. Metamizole sodium)
-The identifier of the drug by NIP (e.g. OGYI-T-07845)
-The filename of the PIL
(e.g. bh_0000018954_20061024094848.doc)
-The drug is sold over the counter or not

This web based tool gives the opportunity to edit/manage the vocabulary of the TTS and the ASR for the operator.

3. Evaluation

The evaluation of the system was executed in two phases: ASR test and system evaluation. In the ASR test we focused on the recognition of the drug names only, in the system evaluation speech of the TTS was tested together with all other components (user friendliness, speed, etc).

The ASR test was organized with six persons (three male and three female aged between 30 and 65). They tested the recognition of altogether 1321 drug names by phone in an office environment with fixed phone. The task was to call up the system, pronounce the drug name and wait for the answer, whether the recognized item was correct or the system failed. In case of wrong recognition the test person pronounced once more the drug name and so on, until four pronunciations. Test results are summarized in Fig 2. In 1281 cases the recognition was successful at once, 29 drug names were recognized only after the second pronunciation, 6 after the third time and 5 of them was not recognized even at the fourth time.

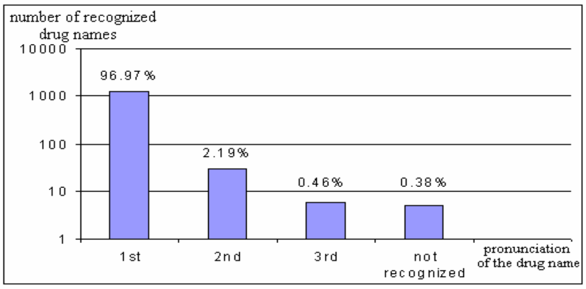

Figure 2: *The recognition scale of 1321 drug names*

System evaluation contained four questions: intelligibility of the voice of TTS, speed of the synthetic speech, user friendliness, speed of getting information. Test persons were asked to evaluate the system features on a 5 level scale (5 = very good, 4 =good, 3 = acceptable, 2 = less acceptable, 1= not good. The test was completed by 57 persons, 15 male and 42 female (aged in three groups: 15 persons under 25 years (A), 33 between 25-60 including 7 blinds (B) and 12 above 62 years (C)). Every test person got a list having 12 drug names. Their task was to call up the system and listen to minimum 2 chapters about the drug. No training was given for them before testing. The results of the system evaluation are shown on Figure 3.

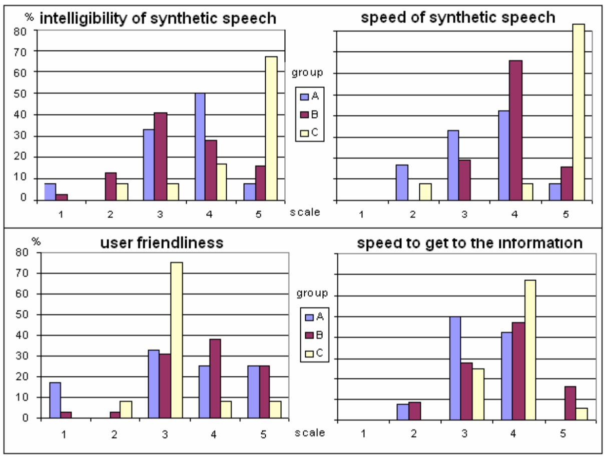

Figure 3: *System evaluation results of Medicine Line*

System evaluation results clearly show the difference between young and elderly generation. Group A found less intelligible the voice of TTS as Group C. As to the speed of synthetic speech Group C found it mostly very good, but young people regarded it too slow. People of group C found user friendliness acceptable, but persons of group A had better opinion. Only question 4 was evaluated by all groups similarly.

The total system evaluation shows the following scale: 1 = 2.25%; 2 = 6.25%; 3 = 30.75%; 4 = 39.25%; 5 = 21.5 %. According to these results the system is mostly evaluated between the acceptable and very good levels.

4. Conclusions

The experiences in using speech technology in this medical field are encouraging. The special word and expression structure of PILs required the construction of a special TTS converter and also the ASR module had to be taught in a special way to recognize drug names. Due to ASR and synthesis features it is clear that pharmaceutical manufacturers have to be involved into the design of such systems. As trademark laws restrict trademarks that are spelled alike, or sound similar to already existing products, the manufacturers of medicines, have to be forced to give distinguishable names for their new products. So, perhaps linguists, phoneticians should be involved when new drug names are designed. The well distinguishable names can be handled by speech technology modules with higher accuracy, and this is the interest of the manufacturers as well. The system is in use since December 2006. The phone number of the Hungarian Medicine Line is +36-1-88-69-490, the home page is: http://www.gyogyszervonal.hu.

5. Acknowledgements

The project was realized together with the National Institute of Pharmacy, Budapest, Hungary. Thanks for their contribution. This research and development was supported by the Hungarian National Office for Research and Technology (GVOP project no. 3.1.1-2004-05-0426).

6. References

[1] Vicsi, K., Velkei, Sz., Szaszák, Gy., Borostyán, G., and Gordos G., "Speech Recognizer for Preparing Medical Reports", Infocommunication, col. LXI.: 14-21, 2006. Budapest, Hungary

[2] Grasso, M.A. "The long term adoption of speech recognition in medical applications", Proceedings of the 16th IEEE Symposium on Computer-Based Medical Systems: 257-262, 2003

[3] Henton, C. "Bitter Pills to Swallow. ASR and TTS have Drug Problems" Int. Journal of Speech Technology. 8: 247-257, 2005.

[4] http://alpha.tmit.bme.hu/speech/hdbMTBA.php

[5] http://alpha.tmit.bme.hu/speech/hdbtesztelen.php

[6] ETSI standard doc., "Speech Processing, Transmission and Quality aspects (STQ); Distributed Speech Recognition", ETSI ES 202-050 v1.1.2.

[7] Mihajlik, P., Tobler, Z., Tüske, Z., Gordos, G. "Evaluation and Optimization of Noise Robust Front-End Technologies for the Automatic Recognition of Hungarian Telephone Speech" in In Proc. International Conference on Speech Communication and Technology, Vol 1, 2677-2680, 2005 Lisbon, Portugal.

[8] Young, S., Ollason, D., Valtchev, V., and Woodland P. The HTK book (for HTK version 3.2.), 2002. March.

[9] Mohri, M., Pereira, F. and Riley M. "Weighted Finite-State Transducers in Speech Recognition", Computer Speech and Language, 16(1):69-88, 2002.

[10] Fegyó, T., Mihajlik, P., Szarvas, M., Tatai, P., Tatai, G., "VOXenter - Intelligent voice enabled call center for Hungarian", In Proc. Eurospeech-2003, 1905-1908.

[11] Olaszy, G., Németh, G., Olaszi, P., Kiss, G., Zainkó, Cs., Gordos, G. "Profivox – a Hungarian TTS System for Telecommunications Applications", Int. J. of Speech Techn., Vol 3-4. Kluwer, 201-215. 2000.