Speech Database Acquisition for Assisted Living Environment Applications

Mihai Dogariu, Horia Cucu*, Andi Buzo, Dragoş Burileanu, Octavian Fratu
Speech and Dialogue Research Laboratory
University Politehnica of Bucharest
Bucharest, Romania
*Corresponding author (e-mail: horia.cucu@upb.ro)

***Abstract*— Home automation has become a subject of increasing interest for both industry and research as there is an increase in the awareness of such systems and their benefits can be easily seen. The new trend is to develop smart homes where commands can be given by speech. This way of communication, besides being the most natural, has the advantage of offering flexibility to the users especially when they have limited motion capabilities. As for widely used languages the state of the art has achieved an important level of performance, little efforts are made with the Romanian language. The main reason for this is the lack of an annotated speech database from real life conditions. This paper focuses on the methodology of acquiring four different speech corpora with various end-user scenarios in mind. The commands corpus is meant to be used in home automation development, the cough corpus is meant to help research in detecting distress situations, the spontaneous speech corpus will aid in distant speech recognition applications and the multi-room, multi-person, multi-language corpus can be used for research in speaker detection and identification. All these were recorded in the context of a completely automated and functional smart home. The small number of such environments available to the public makes these corpora valuable from experimental point of view.**

Keywords—voice-activated home automation; cough detection; distant speaker diarization; distant speech acquisition

I. INTRODUCTION

Speech recognition has gone through significant improvements in the last three decades. A large variety of algorithms and technological breakthroughs have come to the aid of speech recognition. This has opened new domains such as interactive applications in mobile applications or entertainment industry and lately it has helped in the development of distant speech recognition (DSR). With the difficulties posed by speaker orientation towards the microphone, the distance between the speaker and the microphone, environmental noise, room reverberation, echo and many other new variables worth taking into consideration, DSR can be assumed to be more than an expansion of the speech recognition field, up to the point where it deserves an independent topic of discussion in the signal processing field.

As DSR caught an increasing amount of attention from researchers around the globe some research groups started focusing on this kind of applications as the industry developed a need for speech applications in which the speaker is not restricted to using a certain setup for speech recognition (headsets or speaking near a fixed microphone). With the great complexity brought to the task by the numerous factors that must be taken into account and the lack of research in this field, DSR raises challenging problems. A natural result of DSR applications came in the form of equipping an entire house with microphones or microphone arrays in order to capture speech at the same time from more than one room. This helps in the house automation process, also named domotics, which is a field within building automation, specializing in specific automation requirements of private homes [1]. These automation systems often cover utilities such as light controllers, security systems, climate controllers, multimedia entertainment, and many others. The interface with such an automation system is usually a control panel which commands each of the configurable systems present inside the house through a screen and a few buttons.

However, this does not seem to be the most pleasant type of interaction between a person and a house, as it is not convenient to have to be in front of the control panel each time you want to change the state of any household system. Therefore, it is more convenient and natural for the interaction between humans and the control system to be made through voice similar to talking to the control panel instead of operating it in the classical way. Furthermore, voice control provides more flexibility regarding the place where the person controlling the system has to be at the moment when he engages in the human-system interaction. This translates into the fact that the system can be commanded from any room inside the house at any given time, leaving the user with a higher degree of freedom to perform other tasks which imply visual attention or physical labor. Consequently, new applications have been developed having all these advantages in mind and targeting specific user groups which may greatly benefit from having such a system at their own disposal.

Among these people there are those who need to be cared for, namely elders and physically impaired persons, who lack the mobility necessary for performing different tasks which may seem trivial but may pose serious difficulties, e.g. turning the lights off in a different room, maybe even on a different floor, locking all the doors or pulling the blinds. Commanding all these systems by voice can be a great asset for people in need and can significantly improve their quality of life. It can also have a positive impact on the care takers of these people,

978-1-4673-7560-3/15/$31.00 ©2015 IEEE

as it can relieve them from some minor worries and let them focus on more important aspects that cannot be automated. Implementing these facilities is what makes up a so called smart home, which has recently gained an increasing popularity in the housing industry in developed countries.

When it comes to the Romanian language, the efforts in this direction are very limited. The main reason is the lack of an annotated speech database collected from real life conditions. The development of such systems with artificial noise databases has proved to be insufficient [2]. The main difficulty is related to the availability of environments equipped with professional recording devices where controlled experiments can be realized. Therefore the current work will help the advances in this field for the Romanian language.

This paper describes the process of collecting a speech corpus recorded in the premises of a smart home and the set of applications that the corpus is intended to be used in. The corpus has been acquired in a real smart-home, in the DOMUS smart-home of Laboratoire d'Informatique de Grenoble (LIG).

II. Related Work

Even though voice control is undoubtedly an attractive and sometimes even indispensable solution for smart homes, very few projects have included speech recognition in their design [2], [3], [4]. This is probably because of its complexity and the challenges it raises in this particular scenario which are: (1) robustness against noise, (2) distant speech recognition, and (3) the accuracy of keyword spotting. An additional issue with voice control in smart homes is (4) the language dependence it creates, because today's speech recognition systems support only one language and moving to another language requires at least new acoustic, phonetic and language resources which are expensive to obtain. Hence, the solution is not scalable to other languages.

The development of such systems was boosted as soon as real life speech databases were built and used in experiments. Recently, several corpora are developed for studying the distant speech issues within some projects and collaborative programmes. Among them we can mention CHIL [5], AMI [6], REVERB [7], PASCAL-CHIME [8], GRID [9]. The main issues addressed in these initiatives are speech separation, distant speech recognition, and enhancement in noisy environments, based on single or multiple distant microphones. One way of developing such data sets was by using simulations. Such data sets were explored in previous work with promising results [10]. Finally, during the DIRHA project a set of multi-microphone speech corpora was created [11],[12].

In the Romanian language there has been some efforts in acquiring noisy data sets, but these efforts are limited to TV recording which are not suitable for the smart homes distant speech recognition task [13].

III. Targeted Applications

This corpus has been designed with two main projects in mind, namely acquisition of a commands corpus and a coughing corpus. Sometime further in the development of the strategy two more corpora in the premises of the smart-home have been taken into account: one for spontaneous speech and one for multi-language, multi-speaker, multi-room speech. Therefore, a total of 4 different corpora have been recorded, each of them being able to serve more than one application. The applications that were targeted at the start of the project are as follows:

- Commands corpus for automating a smart-home environment to respond to voice commands from anywhere in the house. It was desired to acquire a corpus that would fit a certain set of commands that a user can give to a smart-home's devices and utilities. A certain degree of variability has been also taken into account, managing the commands list in such a way that it covers different ways of expressing the same command and each command was uttered a certain number of times, according to the frequency it is expected to be uttered in a real-case scenario. Therefore, spoken commands were recorded from every speaker, in each room of the house, at varying distances towards the microphones. This would give us information about many aspects influencing the DSR results, namely how the quality of the recording, the distance between the speaker and the recording microphones, speaker's orientation, room reverberation, room echo, indoor and outdoor environmental noise. Having a single person talking in different rooms of the house can be used in an experiment which monitors presence inside the house based on speech solely or combined with other sensors. Another application that was targeted with this corpus is the detection of staccato speech, which is an abrupt utterance, each syllable being enunciated separately, common with people suffering from multiple sclerosis [14], aiding medical research as well.
- Cough corpus for detecting stress situations in which somebody suffers from an acute coughing and the system should automatically detect this and send an alarm message to somebody who can come to help. In order to avoid confusion with normal coughing which can occur from various insignificant reasons the speakers were asked to clear their throat, to perform a short cough and a longer one as well. This can be useful for persons suffering from pulmonary affections or people who may choke and are in need of urgent assistance. Recording this corpus also gives a better understanding of how coughing can be removed from interfering with speech recognition results by being recorded as an additional filler.
- Spontaneous speech corpus for free distant speech recognition that can be useful for transcribing continuous speech and extract information from the context. This can also be mixed with the commands corpus to see how the system behaves when a command appears in the middle of a normal conversation. Based on this corpus, speaker identification or speaker verification systems can also be developed with respect to security issues, which can be a plus for smart-homes.

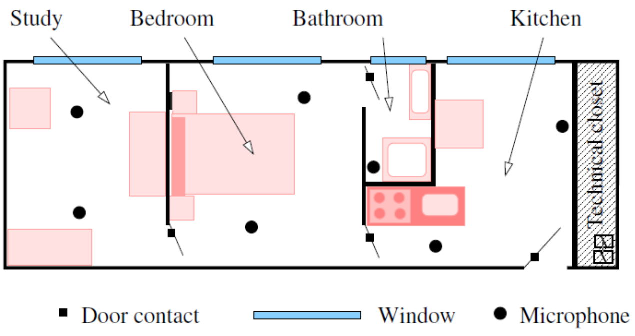

Fig. 1 DOMUS smart-home environment description [15]

- Multi-language, multi-speaker, multi-room corpus for automatically detecting the number of persons in a room. This type of scenarios, commonly referred to as cocktail parties, can be addressed by acoustic source separation techniques. To the already adverse conditions that the cocktail-party situations pose, the multiple rooms through which the speakers can walk, the multiple microphones which are recording at the same time and the multiple languages that were recorded, namely English, French and Romanian have also been added to the list of variables. The targeted applications do not take into account language dependency of any sort, but this can be used for other experiments as well.

All corpora contain both male and female speakers such that gender variability can also be studied. As it can be seen from what has already been presented, the four corpora that were recorded can be the target of numerous applications in the DSR field, depending on the context and on the research objective. Even though the main applications were clearly defined for the database that was created, there are other projects that can benefit from it. It is also worth mentioning that there are not many environments available and already functional like the one in which the recordings have been made, i.e. the DOMUS smart-home owned by LIG.

IV. Environment Description

The DOMUS smart-home of the LIG is fully equipped with sensors and actuators, providing an environment with ambient intelligence and comprises a kitchen, a bedroom, a bathroom and a study room, as presented in Fig. 1. All rooms have the necessary furniture, appliances and they even assure a certain degree of comfort, so the flat is completely usable. The smart-home provides a number of systems, such as blinds, curtains, television, door locks, etc., that can be controlled from an online interface, therefore the commands that the user gives to the house refer to these utilities. Its walls and ceiling are made from audio insulating material in order to keep outer noise to a minimum level. However, the glass windows are not preventing outdoor environmental noise from interfering with the recordings. The house also presents 6 video cameras, two in each room except the bathroom, which from privacy reasons has remained unequipped with such a camera, to monitor people's movement and position throughout the experiment. These footages are used only to verify that the annotation of the speech corpus was correct from the point of view of the room in which the speakers were at the moment of the recording and they are available in the database for any possible further use related to annotation problems. For the audio recording the flat is equipped with 7 radio microphones SENNHEISER ME2 (2 in each room, except the bathroom, where only one is present), set into the ceiling. Out of the 7 microphones there are 4 who give recordings of a better quality, one in each in room. All of them are connected to a National Instrument PCI-6220E multichannel card with the help of which they can record audio signal in real-time, all at the same time [15]. The software that was used to perform the recordings on all the channels at once was StreamHIS [16]. The tool used for transcribing the audio recordings was Transcriber (http://trans.sourceforge.net/). In each session the entire audio signal was captured by all the 7 microphones at the same time and all these recordings are available in the database.

V. Methodology of Database Acquisition

The first step in the corpus acquisition process was to establish the goals regarding what it was desired to have in the end. Therefore, each of the four corpora was recorded starting from some clear scenarios, as it will be described in this section.

The next step was to gather Romanian speakers for the recordings, who established the contact with several Romanian students from Grenoble. They were kind enough to help with the recordings and even contributed to gathering more people to help with the recordings. Their ages range from 22 to 41 years old, as it can be seen in Table I. Some speakers could not record the spontaneous speech or the multi-room, multi-speaker, multi-language part because of the lack of synchronization between their schedule and the time intervals when the environment was available for recording these databases.

The final step was meeting with the volunteers and start the recording itself, which began by explaining the terms and conditions in which the experiments will take place.

TABLE I. Speakers' Demogrpahic Details

Primary speakers				
Speaker id	***Gender***	***Age***	***Height***	***Weight***
S1	M	24	1.87	90
S2	F	23	1.64	54
S3	M	24	1.87	90
S4	F	23	1.69	67
S5	F	23	1.58	53
S6	M	21	1.77	66
S7	F	23	1.7	57
S8	F	41	1.8	84
S9	F	30	1.59	54
S10	F	24	1.58	57
S11	F	21	1.6	56
Auxiliary Speakers				
S12	M	35	N/A	N/A
S13	M	51	N/A	N/A
S14	F	42	N/A	N/A
S15	M	24	N/A	N/A

They were asked to sign a consent form regarding both the audio recording from the microphones and the video recordings from the cameras. They received a signed copy of these agreements and the organizing team another copy, as well.

Before starting each experiment the volunteers were given a complete tour of the house and presented all the rooms and automated systems. Then, they were asked to read the consent agreement aloud, inside the kitchen, for the adaptation of the acoustic model. This short text contains all the phones in Romanian. During the acquisition of the first 2 databases they were alone in the flat and all doors were open.

A. Commands database

For the commands database there has been made a list comprising of all the possible utilities that can be operated by the central control system available in the house that this DSR system will be used in. In order to help people get accustomed to the house, a key word has been introduced as the smart-home's name, i.e. *Casandra*. This is the first word in any command, as it can be seen in Fig. 2, an excerpt from the Finite State Grammar (FSG) that was at the foundation of the commands list. Starting from this FSG, there has been developed a list of possible commands that can be uttered by a speaker wanting to control the automation systems. These commands have been split into 3 categories depending on the frequency with which they are expected to be uttered in a real-case scenario.

The most likely commands (e.g. turning on/off the light) have been uttered by 3 different speakers, the less likely commands (e.g. open/close the blinds) have been uttered by 2 different speakers and the least likely (e.g. start/stop the presence simulator) just by one speaker. Each of the volunteers were given a set of approximately 45 commands interleaved with some sentences which are not related to the ambient intelligence, with a ratio of 7:1, to be able to evaluate some sentences which are not commands in the same environment, with similar conditions. They were asked to repeat this set in each room, except the bathroom, where the set was reduced to a quarter, given the fact that it is less likely to have the same amount of voice activity there. The participants were not given any instructions about what their orientation towards the microphone should be, nor were they restricted regarding their location in the room at the time of speaking. Also, they were not restricted from sitting down at the table, on the bed or sofa and they were not told about the position of the microphones in the house until the end of the experiment unless they would specifically ask for it. They were only told to speak in a normal voice, just as they would talk to another person, with a short break (1-2 seconds) between the commands and to repeat the commands if they felt it is necessary.

All the speakers followed the same scenario, that was provided to them on a sheet of paper, regarding the commands that they should utter and the room in which they should be in at every moment. For every speaker there are available 7 complete recordings, corresponding to each microphone from

```
<electric> = <lumini> | <culoare_lumina> | <dimmer> | <alimentare> |
            <panouri_solare>;

<lumini> = (aprinde|stinge|oprește|pornește|închide|deschide)
            (lumina|[toate] luminile) [[de] la parter|din casă];
<culoare_lumina> = (schimbă|aprinde) (culoarea luminii|lumina)
            (roșie|în roșu|[în] verde|[în] albastru|normală|la
            normal);
<dimmer> = (mărește|crește|micșorează|scade) luminozitatea la
            <numar_zeci> la sută;
<alimentare> = (pornește|reia|oprește) (alimentarea|încărcarea
            bateriei);
<panouri_solare> = care este starea bateriei|este încărcată bateria;
//------------------------------------------------------------------
public <comandaCasa> = casandra (<exterior> | <securitate> |
            <multimedia> | <hvac> | <electric>);
```

Fig. 2. Example of the FSG excerpt that was used to create the commands

the smart-home. In order to annotate the speech, the best recordings have been selected depending on the position where the speaker was when uttering the commands. Consequently, for a better understanding of what was spoken, the audio signal recorded from the microphone from the kitchen while the speaker was in the kitchen was joined with the audio signal from the microphone from the bedroom while the speaker was in the bedroom and with their analogues from the bathroom and the study, as well. To avoid synchronization issues no audio sequence has been cut out in the commands section. The final state of the corpus is detailed in Table II.

B. Coughs database

For the coughs database it was somewhat difficult to differentiate between what can be considered a mild cough during the day and what can be considered a serious, critical cough that can threaten one's health. Therefore, some aspects had to be taken into account, namely to make separate coughing sessions each of them consisting of one voice clearing, one short cough and one long cough. Having noticed that it is exhausting to perform these coughs repeatedly without taking any break some restrictions have been imposed in this way and resulted in interleaving these sessions with the commands in order for the speakers not to lose their voice or get exhausted. These restrictions have been followed in most of the cases, with some minor flaws where the speaker would forget to perform a long cough or make a short cough instead of a long cough. However, this exceptions have been treated and annotated accordingly. One aspect that was noticed and may be useful in further research is that long coughs are very similar to a repeated short cough. That may be dependent on the fact that the human lungs can exhale only so much air as to perform a short cough, therefore in longer coughs this process is repeated several times. Unfortunately, no natural coughs could have been recorded as all the volunteers were healthy, but what they managed to simulate is close to the real-case scenarios. In the end, each speaker recorded 2 coughing sessions in each room, except in the bathroom where only one coughing session was recorded with plenty of time between them so that the speakers could rest their voice. For annotation, the same technique as in the previous case was applied. Details about what was recorded can be seen in Table II.

Fig. 3. Video recording example from the cameras inside the smart-home

C. Spontaneous speech database

For the spontaneous speech database there has been made a list of questions that were asked to each speaker that attended this scenario in order to provide as much free speech as possible. The list of questions was designed in such a way that it would be possible for everyone to offer long and complex answers. The questions were not intrusive or too personal in any way that could make the speakers uncomfortable. At first, there was an attempt to ask the questions via laptop messaging but that was time-consuming and the speakers wouldn't offer long enough answers as the communication was somewhat artificial. Therefore, one of the organizers sat with the volunteer in the study and asked these questions and kept his interventions to a minimum. This was better as he could also ask additional questions based on the answers that the participants provided so the dialogue was richer and longer. Details regarding this corpus can be found in Table II.

D. Multi-room, multi-speakers, multi-language database

For the multi-room, multi-speakers, multi-language database some scenarios have been designed so they would guarantee the presence of different number of persons in each room, some transitions from one room to another, both with the doors open and closed, acquiring speech in multiple languages from speakers of both genders. These recordings were performed with 4 people inside the flat, divided into 2 groups. To make sure that everything runs smoothly one of the organizers was the leader of the experiment, being also part of one of the two groups. The scenario went as follows: all 4 people started from the kitchen and were discussing for 1-2 minutes, with the door between the kitchen and the bedroom closed. Then, the leader would take another person with him and go into the study, closing the door to the kitchen behind them. From technical issues the door between the bedroom and the study could not be closed so it remained open during all the experiments. Once the two groups were separated in the kitchen and study, respectively, they each had conversations for another minute. After that, the leader asked someone from the kitchen to open the door and the conversations continued in each room for another minute. This was done to have some inter-room interference as well, to have it available for future research with as many types of speech as possible. Then the leader went in the bedroom and simulated talking at the phone with a friend, while the group from the kitchen continued the conversation and the one remaining in the study would remain quiet. Then, the leader would start calling people in the bedroom to join a conversation starting with the person that remained alone in the study, then the persons from the kitchen, leaving the door to the kitchen open. After each person came into the bedroom the group that was increasing in number would make a short conversation for about 1 minute until the leader would call another person. Some discussion topics have been chosen apriori to the experiment to avoid situations in which people would not have what to talk about, given the fact that they barely met a few minutes before the experiment. These subjects have been printed out and given to the volunteers. Fig. 3 presents a moment from this experiment, in which there were 3 people inside the bedroom recorded by the video cameras.

TABLE II. RECORDED DATABASES DETAILS

Corpus name	*Number of involved speakers*	*Total recordings duration*	*Observations*
Commands	11	00:58:05	1764 commands were recorded
Coughs	11	00:06:10	240 coughs were recorded, split in 3 roughly equal proportions between throat clearing, short cough and long cough
Spontaneous Speech	8	01:53:50	Strongly dependent on speakers' willingness to offer rich answers; only 2 speakers have been adnotated
Multi-room, multi-person, multi-language	12	00:53:46	Strongly dependent on speakers' willingness to offer rich answers

Having all the persons talking for some short time frames was also challenging because this also depends on one's personality so the leader had to ask questions individually to everybody in order to gather speech from every participant. One session of multi-speaker, multi-room, multi-language recording lasted roughly 10 minutes and there were no constrains regarding position, loudness or conversation topic. The purpose of this database was to acquire speech signal from more persons and from more rooms at the same time. Having more than one language being used during the recordings should not have a strong impact on the results. Table II gives information about what was recorded in all the 4 corpora.

VI. CONCLUSIONS

With the smart-home industry growing it is clear that speech will play a vital role in its implementation. However, the current state of development does not allow for many tests to be performed in such an environment, but in the near future there will be more such intelligent buildings which could support further research in this domain. This paper presents the acquisition of 4 different databases that can have multiple uses each, namely the commands corpus, coughs corpus, spontaneous speech corpus and the multi-room, multi-speaker, multi-language corpus. All these databases have been recorded in a fully equipped smart-home. The entire process of recording these corpora has been presented, along with the resulting data. However, the obtained data is not sufficient for implementing a reliable system due to the low number of speakers that were involved, but it can serve to several systems in the evaluation process, as they are suited to a large set of applications.

ACKNOWLEDGMENT

The speech database was recorded in Grenoble, France, under the supervision and guidance of Michel Vacher and François Portet of the GETALP team of LIG to whom we wish to express our gratitude. All the speakers have volunteered for these recordings and we would like to thank them, as well, for being so helpful and cooperative.

This work has been partially funded by European Commission through the FP7 IP project no. 610658/2013 ”eWALL for Active Long Living - eWALL”, partially by UEFISCDI Romania under grant no. 262EU „eWALL” support project and partially by the PN II Programme "Partnerships in priority areas" of MEN - UEFISCDI, through project no. 32/2014.

REFERENCES

[1] A. Badii and J. Boudy, “CompanionAble - integrated cognitive assistive & domotic companion robotic systems for ability & security,” in 1st Congres of the Société Française des Technologies pour l’Autonomie et de Gérontechnologie (SFTAG’09), Troyes, 2009, pp. 18–20.

[2] A. Vovos, B. Kladis, and N. Fakotakis, “Speech operated smarthome control system for users with special needs,” Proc. of Interspeech, 2005, pp. 193–196.

[3] M. Hamill, V. Young, J. Boger, and A. Mihailidis, “Development of an automated speech recognition interface for personal emergency response systems,” Journal of NeuroEngineering and Rehabilitation, vol. 6, 2009.

[4] B. Lecouteux, M. Vacher, F. Portet, "Distant Speech Recognition in a Smart Home: Comparison of Several Multisource ASRs in Realistic Conditions", Proc. of Interspeech, pp. 2273-2276, 2011.

[5] D. Mostefa et al., “The CHIL Audiovisual Corpus for Lecture and Meeting Analysis inside Smart Rooms,” Language resources and evaluation, vol. 41, no. 3, pp. 389–407, 01/2008 2007.

[6] J. Carletta et al., “The AMI Meeting Corpus: A Pre-announcement,” Proc. of the Second International Conference on Machine Learning for Multimodal Interaction, 2006.

[7] K. Kinoshita et al., “The REVERB challenge: A Common Evaluation Framework for Dereverberation and Recognition of Reverberant Speech,” in IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), 2013, pp. 1–4.

[8] J. Barker, E. Vincent, N. Ma, H. Christensen, and P. Green, “The PASCAL CHiME speech separation and recognition challenge.,” Computer Speech and Language, vol. 27, no. 3, pp. 621–633, 2013.

[9] M. Cooke, J. Barker, S. Cunningham, and X. Shao, “An audio-visual corpus for speech perception and automatic speech re-cognition,” The Journal of the Acoustical Society of America, vol. 120, no. 5, pp. 2421–2424, November 2006.

[10] M. Matassoni, M. Omologo, D. Giuliani, and P. Svaizer, “HMM Training with Contaminated Speech Material for Distant-Talking Speech Recognition,” Computer Speech and Language, 2002, pp. 205-223.

[11] L. Cristoforetti, M. Ravanelli, M. Omologo, A. Sosi, A. Abad, M. Hagmuller, P. Maragos, “The DIRHA simulated corpus,” Proc. of LREC, 2014.

[12] M. Matassoni, R. Astudillo, A. Katsamanis, and M. Ravanelli, “The DIRHA-GRID corpus: baseline and tools for multi-room distant speech recognition using distributed microphones,” Proc. of Interspeech, 2014, pp 1613-1617.

[13] H. Cucu, A. Buzo, C. Burileanu, “Unsupervised acoustic model training using multiple seed ASR systems,“ Proc. of the 4th International Workshop on Spoken Language Technologies for Under-resourced Languages (SLTU), St. Petersburg, Russia, 2014, pp. 124-130.

[14] Medical Dictionary for the Health Professions and Nursing © Farlex 2012.

[15] M. Vacher, B. Lecouteux, P. Chahuara, F. Portet, B. Meillon, et al.. “The Sweet-Home speech and multimodal corpus for home automation interaction,“ The 9th edition of the Language Resources and Evaluation Conference (LREC), May 2014, Reykjavik, Iceland. pp.4499-4506.

[16] B. Lecouteux, M. Vacher, and F. Portet, “Distant speech recognition for home automation: Preliminary experimental results in a smart home,“ In IEEE SPED 2011, pages 41–50, Brasov, Romania, May 18-2.