Proc. 14th Int. Conf. on Global Research and Education, Inter-Academia 2015
JJAP Conf. Proc. **4** (2016) 011614
©2016 The Japan Society of Applied Physics

Adaptive Speech Recognition Framework for Dysarthric Patients

Gabriella Simon-Nagy[1*], and Annamária R. Várkonyi-Kóczy[2,3,4*]

[1]Doctoral School of Applied Informatics and Applied Mathematics, Óbuda University, Budapest, Hungary
[2]Institute of Mechatronics and Vehicle Engineering, Óbuda University, Budapest, Hungary
[3]Integrated Intelligent Systems Japanese-Hungarian Laboratory
[4]Department of Mathematics and Informatics, J. Selye University, Komarno, Slovakia

E-mail: nagy.gabriella@nik.uni-obuda.hu, varkonyi-koczy@uni-obuda.hu

(Received September 27, 2015)

Dysarthria is a speech disorder that mostly occurs as a symptom of neurodegenerative and other neuromuscular diseases. The speech of patients with dysarthria becomes distorted, the articulation of phonemes (especially that of consonants) is poor, the intelligibility and naturalness are impaired. Because dysarthria is progressive (similarly to the other symptoms of the main disease), patients may have difficulties using speech-controlled Ambient Assisted Living systems that could be a great help for them in daily life. In this paper, an adaptive speech recognition framework is introduced that is able to handle gradually occurring changes in the speech quality of the user. The presented technique can adapt to these changes while the speech interpretation accuracy of the system will not decrease, even in cases of noisy or incorrect training data.

1. Introduction

1.1 Smart homes for motion disabled persons

Progressive neuromuscular diseases such as multiple sclerosis (MS) or amyotrophic lateral sclerosis (ALS) cause severe motion disability as years go by. Speech-controlled Ambient Assisted Living (AAL) or smart home environments are able to provide part of the necessary help instead of a human caregiver. However, there is a significant challenge in designing a speech-recognition solution for these patients because of their speech-impairment (dysarthria): poor articulation, impairment in phonation, and prosody. The quality of speech deteriorates with the progression of the disease; the associated symptoms (slowly but constantly) change (Refs. 1 and 2).

Therefore, the usual speech recognition approaches (speaker-independent model trained with speech samples of healthy persons; model pre-trained with samples from the user) may not be reliable enough in this case. The recognizer must have a means of adapting to the changing parameters of the speech of the user.

Further, for an AAL system supporting disabled persons, additional requirements must be considered to meet the needs of the user. If a smart home system misinterprets a command and e.g. accidentally switches the air conditioning off, a healthy person can correct it by just pushing a button. A motion disabled user may not be able to do that and mistakes like this can potentially cause dangerous situations. It is more acceptable to ask for confirmation of the command (or the user having to repeat it) than to execute unwanted tasks. Of course, alarm situations and distress signals are exceptions. In such cases, execution should not be delayed and false positives are more acceptable than the incorrect rejection of the command, especially if there is a way to cancel the call if unwanted. The assistance call has to be sent as quickly as possible even if later it proves to be a false alarm.

1.2 Test environment

In the following, a new adaptive speech recognition framework is presented, designed for an existing AAL system. The system has been developed for an elderly woman with multiple sclerosis. She is not able to move her legs or hands anymore (not even a finger). She can move only her head to a limited extent. Although, she is still able to speak in a relatively comprehensible manner but already shows several symptoms of dysarthria which has limited the possible solutions of voice control to be used in the AAL system.

The control software runs on a regular PC (CPU: Intel Core i5 2x1.6 GHz, 4 GB DDR3 RAM, Windows 7 SP1 OS, integrated sound card, simple microphone with frequency range 20-16000 Hz, 50dB S/N ratio). The control system executes the speech recognizer algorithm and also manages the tasks initiated by the spoken commands. The hardware items are controlled by the main controller that connects to the PC by USB. The main controller communicates with the peripheral hardware items using the ZigBee wireless standard.

The online radio application and the Skype run on the same PC as the control software. (Skype does not support this kind of controlling anymore; therefore it is an older version of the software.) The PC can be remotely accessed using the TeamViewer software, but the function is password protected, thus only the authorized administrators can log in.

The commands of the system are organized in a menu hierarchy to help the user in remembering them. Being in the main menu the user can ask for the submenus (like bed control, radio control, television control, etc.), as well as for some distinguished, single-function commands (to tell the temperature or the time, or switch the lights on/off). There are also commands which are always (at any point of the menu tree) available (like emergency calls, nurse call, and the quit instruction).

The implemented system applies anytime speech recognition over a limited set of commands organized in a menu structure. Interruptible analyzer modules calculate the probabilities of the heard instruction matching the recorded samples of the pre-defined commands based on the analysis/matching of different speech parameters. Then, the commands with matching probability below a given acceptance threshold, are ruled out. The modules look for various speech parameters which are easily recognizable in the given non-optimal acoustic environment and are relatively unimpaired by dysarthric symptoms. The recognizer was not capable to adapt its models before, but the application of the proposed rules and structure would qualify it for following the speech changes of the user.

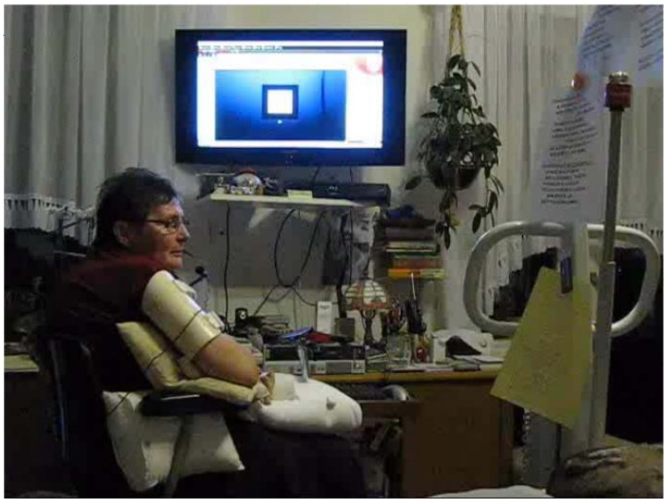

Fig. 1. Part of the implemented system.

2. Adapting recognition methods

2.1 Adaptive methods for speech recognition tasks

Adaptive methods are widely used in speech recognition, for different purposes such as noise adaptation (Refs. 3 and 4), orientation to specific accents of a language (Ref. 5), or choosing useful training data for labeling from a large corpus of unlabeled speech samples (Ref. 6). The adapting system can be one of the well-known soft computing approaches like neural networks (Ref. 7) or fuzzy filters (Ref. 4).

However, a continuous adaptation method running during live operation may pose a risk to recognition accuracy especially because of background noises in the living environment containing voices of other people or TV shows. Since the smart home of a motion disabled person has to meet the requirements of a safety-critical system, non-decreasing recognition accuracy must be guaranteed if possible.

2.2 Rules for adapting recognition methods

In the following, a number of rules suitable for any speech recognition method with a continuous adaptation capability are discussed. The requirements that must be observed are:

- **The recognizer should adapt to the deterioration of speech quality during live operation.**

  The AAL system is in use constantly, therefore adaptation to slow changing of speech should happen without disturbing or interrupting normal operation.

- **The recognition accuracy should not decrease due to the adaptation of the recognizer.**

  The adaptation process should be robust against noise (i.e. teaching the system with noisy speech samples should not have a negative impact on accuracy). However, accuracy may decrease temporarily due to the changes in the speech of the user.

- **The system should have means of detecting faulty categorization.**

  Detecting categorization errors is not trivial in such systems because of the difficulties of the user in correcting unwanted behavior. There should be a method of indicating a mistake orally (by a special, easily detectable command) or any other way the user is capable of.

2.3 Multiple recognizers

The above requirements can be achieved by using a set of recognizers instead of just one. The recognizer may be an artificial neural network, a set of fuzzy filters, or any other soft computing solutions. Several criteria must be applied when building the recognition system:

1. Recognizers must be independent of each other, so that the fault of one does not affect the others.
2. Different adaptation policies (and/or teaching samples) can prove useful in creating diversity among the learned parameters of the recognizers, and to promote noise-tolerance.
3. One recognizer (preferably an instance of the most accurate recognizer) must be excluded from the adaptation process and remain unchanged to prevent decrease in accuracy. The excluded instance can be changed later to a more accurate one.
4. Adapting recognizers may or may not participate in constructing the output of the system.

Such systems can be built in different ways. An example is shown in Fig.2.

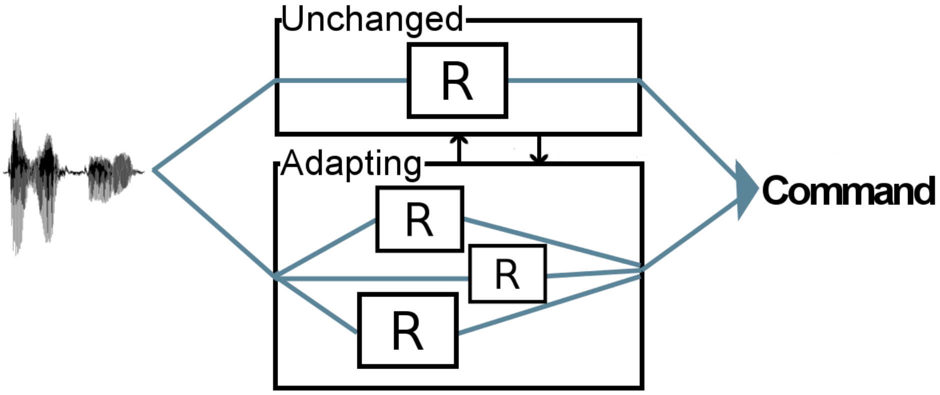

Fig. 2. Possible structure of an adapting speech-recognition system.

The solutions given by the individual recognizers have to be combined to a single output. Several decision-making methods can be used to achieve this, however, not all the potential solutions should be taken into account with equal weights. It is recommended for the weights to be proportional to the reliability of the recognizer, which can be calculated based on the accuracy of a given number of recognitions in the past.

An example of such combination is the following: if every recognizer gives a command as a potential solution, then a score can be determined for every command as the sum of the reliability scores of all the recognizers that the given command is suggested by. This method has the advantage of quick and simple calculation, and a maximum search on the scores list can determine the combined output.

After calculating the output, adapting recognizers must be taught according to their adaptation policies (each recognizer should have a different policy to promote robustness), and reliability scores must also be recalculated. If the reliability score of a recognizer drops below a given threshold, it should be deleted and replaced by a copy of the most accurate recognizer instance.

3. Conclusion

In this paper, an adaptive speech recognition framework is proposed. The most important requirements are also formulated to be able to provide an adaptive, safe, and reliable system for motion disabled persons suffering from dysarthria. Although testing has begun in a test environment with a single user, further extensive testing of the concept is needed on speech samples collected from patients for years.

Acknowledgment

This work was supported by the Hungarian National Scientific Fund (OTKA 105846).

References

[1] L. Hartelius, B. Runmarker, and O. Andersen, Folia Phoniatrica et Logopaedica: Official Organ of the International Association of Logopedics and Phoniatrics, p. 160 (2000)

[2] B. Tomik, and R.J. Guiloff, Amyotrophic Lateral Sclerosis, Vol. 11, p. 4 (2010)

[3] O. Kalinli et al., IEEE Trans. Audio, Speech, and Language Processing, Vol. 18, Issue 8, p. 1889 (2010)

[4] C. Juang and C. Lin, IEEE Trans. Fuzzy Systems, Vol. 9, Issue 1, p. 139 (2001)

[5] H. Wang, L. Wang, and X. Liu: Proc. 4th IEEE Int. Conf. on Information Science and Technology (ICIST), Shenzhen, 2014.

[6] G. Riccardi and D. Hakkani-Tur, IEEE Trans. Speech and Audio Processing, Vol. 13, Issue 4, p. 504 (2005)

[7] S. Xue et al., IEEE/ACM Trans. Audio, Speech, and Language Processing, Vol. 22, Issue 12, p. 1713 (2014)