SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 5 NO 4, 2017
ISSN 2054-7390

Building An Automatic Speech Recognition System for Home Automation

[1]Mohamed Aboulkhir, [1]Samira Khoulji, [2]Reda Jourani, ML Kerkeb
[1]Spatial Remote Sensing -Signal Processing Applied Maths &Computer Science Decision Support Laboratory, Abdelmalek Essaadi University, National School of Applied Sciences, 93000,Tetouan,Morocco.
[2]Polydisciplinary Faculty of Tetouan, Abdelmalek Essaadi University, Tetouan, Morocco.
amohamed.aboulkhir@gmail.com, bkhouljisamira@gmail.com,creda.jourani@yahoo.fr, kerkebml@gmail.com

ABSTRACT

This paper presents a study on automatic speech recognition (ASR) systems applied to home automation. So a detailed study of the architecture of speech recognition systems was carried out. The objective is to select a speech recognition software that must operate in remote speech conditions and in a noisy environment. The proposed system is using an ASR toolkit called Kaldi, which must communicate as an open platform communication (OPC) client developed in C++, with any home automation system. The latter behaves like an OPC server.

Keywords : Speech recognition, acoustic model, language model, HMM, n-gram, domotics,Kaldi.

1 Introduction

Speech is the most natural mode of communication. Through it, we can give voice to our thoughts. We can use it to express opinions, ideas, feelings, desires or to exchange, transmit, request information. And today we do not just use it to communicate with other humans, but also with machines.

Speech recognition is the technique that allows the analysis of sounds picked up by a microphone to transcribe them into a series of words that can be used by machines. Since its appearance in the 1950, automatic speech recognition has been constantly improved. Today the applications of speech recognition are very diverse and each system has its own architecture and mode of operation. The wider the field of application, the greater the recognition models must be (in order to understand spontaneous discourses and the diversity of the speakers)[1].

In our study, we will be interested in the automatic recognition of speech applied to home automation. Indeed, it is the goal pursued by the smart home which is a residence equipped with computer technology to assist its inhabitants in the various situations of the domestic life as well in terms of comfort as that of security. Automatic Speech Recognition could be an essential contribution to the detection of abnormal situations, which is an essential part of a home surveillance system [2].

The paper is organized as follows: we start by describing the components of a speech recognition system. This is followed by describing components of many speech recognition software and testing their

DOI: 10.14738/tmlai.54.3190
Publication Date: 15th August, 2017
URL: http://dx.doi.org/10.14738/tmlai.54.3190

performance in order to choose the one, which will subsequently be integrated into a home automation system.

2 Automatic Speech Recognition

2.1 Overview

A speech recognition system is intended to associate a sequence of words with a sequence of acoustic observations. Thus, from the sequence of acoustic observations X, this system searches for the sequence of words Ŵ which maximizes the probability P (W | X)

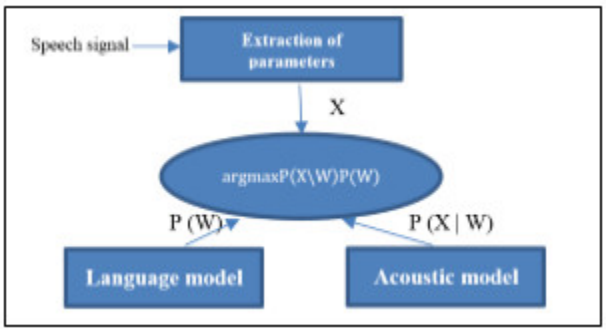


Figure. 1 Schematics of the operation of a speech recognition system

which is the probability of emission of W knowing X[3]. The sequence of words Ŵ must then maximize equation:

$$\hat{W} = argmaxP\left(\frac{W}{X}\right) \quad (1)$$

Applying the Bayes rule, we obtain the formula

$$\hat{W} = argmax\frac{P(X\backslash W)P(W)}{P(X)} \quad (2)$$

Since $P(X)$ is constant, then:

$$\hat{W} = argmaxP(X\backslash W)P(W) \quad (3)$$

Two types of probabilistic models are used to search for the most probable sequence of words: an acoustic model that provides the value of P (X | W), and a language model that provides the value of P (W). Fig. 1 shows a general schematic diagram of the operation of an automatic speech recognition system[4].

2.2 Feature Extraction

As can be seen in Fig. 1, the speech signal cannot be directly transformed into hypotheses of word sequences. The extraction of its parameters is an important step since it must determine the relevant characteristics of the signal. This extraction can be done using multiple techniques, the most well-known ones being parametric analysis, using the LPC (Linear Predictive Coding) method, the cepstral analysis, with for example the MFCC (Mel-scale Frequency Cepstral Coefficients), or the PLP (Perceptual Linear Prediction) technique. These different methods make it possible to extract characteristic coefficients for each frame. This extraction then makes it possible to obtain the sequence of acoustic observations X [5].

2.3 Acoustic Model

The acoustic model is a statistical model that estimates the probability that a phoneme has generated a certain sequence of acoustic parameters. A wide variety of acoustic parameter sequences are observed for each phoneme due to all variations related to speaker diversity, age, gender, dialect, state of health, emotional state. The most widely used methods for acoustic modeling are models based on hidden Markov models (HMM) or deep neuron networks (DNN)[6].

2.4 Language Model

Language models are processes that estimate the probabilities of the different word sequences P (W). These models are used to memorize sequences of words from a textual corpus of learning. In the context of speech recognition, language models serve to guide and constrain research among alternative word hypotheses[7]. The most commonly used language models are n-gram models

2.5 Evaluation of speech recognition systems

In order to evaluate several speech recognition systems, they should be compared on the same test data. Conventionally, these systems are evaluated in terms of word-error rates[8]. The WER takes into account the errors of:

- Substitution: recognized word in place of a word of manual transcription.
- Insertion: Recognized word inserted in relation to the reference transcription.
- Deletion: word of forgotten reference in the hypothesis provided by the speech recognition system.

The WER is expressed by the formula:

$$WER = \frac{substitutions+insertions+suppressions}{\text{number of words in the reference}} \quad (4)$$

3 Automatic speech recognition tool

There are several open-source software for automatic speech recognition (ASR). Notable among these are HTK, Julius (both written in C), Sphinx-4 (written in Java), RWTH ASR toolkit and Kaldi (both written in C++)[9].

In the first phase, we selected an ASR software with the characteristics adapted to the construction of our system, or the least satisfactory to the best of all these needs. After we will describe the features of automatic speech recognition supported by this software.

3.1 ASR Software: Kaldi

To advance our study we choose the voice recognition software called Kaldi. It is an open-source toolkit for speech recognition written in C++ and licensed under the Apache License v2.0.[10]. This choice is motivated by:

- Kaldi have modern and flexible code written in C++ that is easy to understand, modify and extend.
- Open license: The code is licensed under Apache v2.0, which is one of the least restrictive licenses available.

- Extensible design: The algorithms are developed in the most generic form possible. This will allow us to easily integrate our home automation.
- Extensive linear algebra Support: it include a matrixlibrary that wraps standard routines.
- Complete recipes: it make available complete recipes for building speech recognition systems that work from widely available databases.
- Performance: Kaldi outperforms all the other recognition toolkits[11].

3.2 Overview of Kaldi

Kaldi is a speech recognition toolkit consisting of a library, command lineprograms and scripts for acoustic modelling [12].The architecture of Kaldi, as described in Figure 2, consists of the following modules:

- External Libraries: Kaldi depends on two external libraries that are also freely available. One is OpenFst for the finite-state framework (FST), and the other is numerical algebra libraries such as BLAS "Basic Linear Algebra Subroutines" and LAPACK "Linear Algebra PACKage"[13].
- Kaldi C++ Library: It contains all the functionalities and different modes of a speech recognition system developed by C ++, which are then called from a scripting language for building and running a speech recognizer.
- Kaldi C++ Executables Scripts

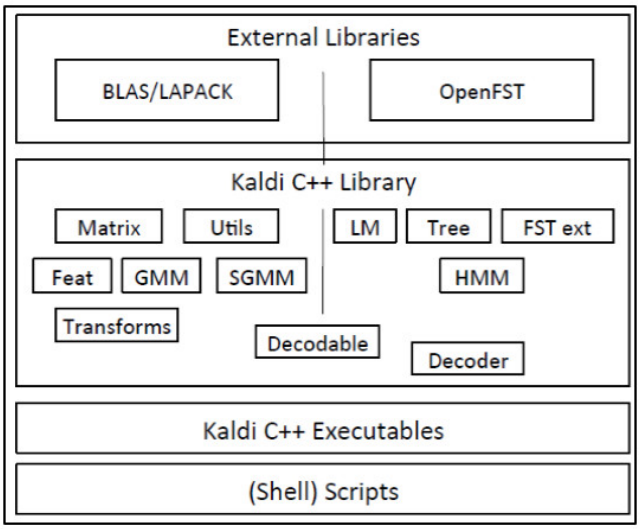


Figure. 2 – Kaldi architecture

As described in the second part of this document, a speech recognition system consists of three modules: Parameter extraction, acoustic model, language model. The Kaldi library integrates these three elements and proposes the following codes:

- For the extraction of parameters: the Kaldi code aims at creating standard functions of MFCC and PLP, setting reasonable default values, but leaving people the opportunity to change the values.
- For the acoustic model: Kaldi supports conventional acoustic models such as GMM, SGMM, HMM and DNN, it is also extensible to new types of models.

```
HRESULT hr;
hr = CLSIDFromProgID(lpwSeverName, &clsid);
hr = CoCreateInstance(clsid, NULL, CLSCTX_ALL, IID_IUnknown, (LPVOID *)&pUnkn);
// Get connection-pointer
hr = pUnkn->QuervInterface(IID_IOPCServer, (LPVOID*)&m_pOpcServer);
```

- For the language models: Kaldi allows to use any language model that can be represented as a FST. Therefore,it supports the most used n-gram model[14].

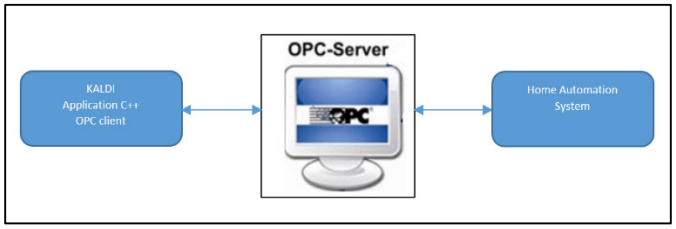
4 Kaldi for Domotics

The aim of our work is to integrate our chosen speech recognition software Kaldi, to a home automation system. For this, we will first propose an architecture of integration, which will be the object of several tests in order to qualify the performances of our model.

4.1 Integration Architecture

In a first step, we propose a communication architecture between Kaldi and a home automation system. Our proposal was based on an architecture based on the OPC client / server technique (Figure 3)[15]. This choice is motivated by:

- The C ++ language supports OPC. Thus, Kaldi can be configured as an OPC client [16].
- The client / server communication can be integrated with the most home automation systems[17].




```
// Write by IOPCSyncIO.
hr = m_pOPCGroup->QueryInterface(IID_IOPCSyncIO, (LPVOID*)&pOPCSync);
```

Figure. 3 – Integration architecture

4.2 OPC Routines for kaldi

We will detail in this section the steps developed on Kaldi for the implementation of the client / server OPC communication between our ASR software (called client) and home automation system through its OPC server.The first step is to connect to OPC Server;the following code will allow this connection:

The second step is to Create OPC Group Object and Add Tags

The last step is to configure read and write to and from an OPC server. In our case, we will only configure the write:

```
// Add OPCGroup
HRESULT hr = m_pOpcServer->AddGroup(GroupName.AllocSysString(), Active,
                    UpdateRate, (OPCHANDLE)pNewGroup, &Bias, &Deadband,
                    LocaleID, &(pNewGroup->m_hServerHandle), &Rate,
                    IID_IOPCGroupStateMgt, (LPUNKNOWN*)&pInterface);

// Add OPCItems
OPCITEMDEF* idef = new OPCITEMDEF[nPoints];
for(DWORD i=0; i<nPoints; i++)
{
  CItemObj* pItemObj = TempList->GetAt(pos);
  CString csName = pItemObj->m_Name;
  idef[i].szItemID = csName.AllocSysString();
  idef[i].dwBlobSize = 0;
  idef[i].pBlob = NULL;
  idef[i].bActive = TRUE;
  idef[i].hClient = (OPCHANDLE)pItemObj->m_hClientHandle;
  idef[i].szAccessPath = AccessPath.AllocSysString();
  idef[i].vtRequestedDataType = VT_EMPTY;
  TempList->GetNext(pos);
}
hr = m_pOPCGroup->QueryInterface(IID_IOPCItemMgt, (LPVOID*)&m_pOPCItem);
hr = m_pOPCItem->AddItems(nPoints, idef, &pResults, &pErrors);
```

5 Conclusion

After the presentation of the state of the art of automatic speech recognition systems,we described the design of Kaldi, a free and open-source speech recognition toolkit. It supports a wide range of methods for extracting parameters, acoustic models and language models. We were also able to configure Kaldi as an OPC client in order to be able to integrate it into a home automation system through its OPC server.Future work will concentrate on implementing this integration in order to test the robustness of our system.

REFERENCES

[1] Allauzen A. et Gauvain J.-L., Construction automatique du vocabulaire d'un système de transcription, dans Journées d'Étude sur le Parole (JEP)

[2] Michel Vacher. Analyse sonore et multimodale dans le domaine de l'assistance à domicile. Intelligence artificielle [cs.AI]. Université de Grenoble, 2011.

[3] Richard Dufour. Transcription automatique de la parole spontanée. Informatique [cs]. Université du Maine, 2010.

[4] Mohamed Bouallegue. L'analyse factorielle pour la modélisation acoustique des systèmes dereconnaissance de la parole. Autre [cs.OH]. Université d'Avignon, 2013.

[5] Insect sound recognition based on mfcc and pnn. In Multimedia and Signal Processing (CMSP), 2011 International Conference IEEE

[6] Fethi Bougares. Attelage de systèmes de transcription automatique de la parole. Ordinateur et societé [cs.CY]. Université du Maine, 2012.

[7] Panagiota Karanasou. Phonemic variability and confusability in pronunciation modeling forautomatic speech recognition. Other [cs.OH]. Université Paris Sud - Paris XI, 2013.

[8] Ngoc-Tien Le, Christophe Servan, Benjamin Lecouteux, Laurent Besacier. Better Evaluatio of ASR in Speech Translation Context Using Word Embeddings. Interspeech 2016.

[9] AMAN F., VACHER M., PORTET F., DUCLOT W. & LECOUTEUX B. (2016). CirdoX : an On/Off-line Multisource Speech and Sound Analysis Software. In LREC 2016.

[10] Madikeri, S., Dey, S., Motlicek, P., & Ferras, M. (2016). Implementation of the standard i-vector system for the kaldi speech recognition toolkit (No. EPFL-REPORT-223041). Idiap.

[11] Gaida, C., Lange, P., Petrick, R., Proba, P., Malatawy, A., & Suendermann-Oeft, D. (2014). Comparing open-source speech recognition toolkits. Tech. Rep., DHBW Stuttgart.

[12] The Kaldi Speech Recognition Toolkit, Povey Daniel, Ghoshal Arnab, Boulianne, GillesBurget, LukasGlembek, OndrejGoel, Nagendra, Hannemann, MirkoMotlicek Petr, Qian Yanmin, Schwarz Petr, Silovsky Jan, Stemmer Georg and Vesely Karel, Idiap-RR-04-2012

[13] Povey, D., Hannemann, M., Boulianne, G., Burget, L., Ghoshal, A., Janda, M., ... & Riedhammer, K. (2012, March). Generating exact lattices in the WFST framework. In Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference. IEEE.

[14] C. Allauzen, M. Riley, J. Schalkwyk, W. Skut, and M. Mohri, “OpenFst:a general and efficient weighted finite-state transducer library,” in Proc.CIAA, 2007.

[15] Olivier Passalacqua, Eric Benoit, Marc-Philippe Huget, Patrice Moreaux. INTEGRATINGOPC DATA INTO GSN INFRASTRUCTURES. IADIS International Conference APPLIEDCOMPUTING 2008

[16] Zheng, L., & Nakagawa, H. (2002, August). OPC (OLE for process control) specification and its developments. In SICE 2002. Proceedings of the 41st SICE Annual Conference (Vol. 2, pp. 917-920). IEEE.

[17] Topalis, E., Orphanos, G., Koubias, S., & Papadopoulos, G. (2000). A generic network management architecture targeted to support home automation networks and home internet connectivity. IEEE Transactions on Consumer Electronics, 46(1), 44-51.

