International Journal of Engineering & Technology, *7 (3.4) (2018)* 177-180

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A New Digital Solution Helps Automatic Voice Recognition

M. ABOULKHIR[1]*, S. BOUREKKADI[2], S. KHOULJI[3] , K. SLIMANI[4], M. L. KERKEB[5]

[1,3]ERISI, AbdelmalekEssaadi University, National School of Applied Sciences, Tetouan,Morocco.
[2]RL Management Sciences of Organizations, the national school of commerce and management, IbnTofailUniversity , Morocco
[4]Faculty of science, IbnTofail University, Kenitra, Morocco
[5]Polydisciplinary Faculty of Larache, Morocco
**Corresponding author E-mail:Mohamed.aboulkhir@gmail.com*

Abstract

This scientific work concerning an examination on automatic speech recognition (ASR) frameworks connected with the home automation and to express the importance of this academic work, an itemized investigation of the engineering of speech recognition frameworks was completed. Our goal in Information Systems Engineering Research Group ofAbdelmalekEssaadi University is to choose a speech recognition programming that must work in remote speech conditions and in a rowdy area.
The proposed framework is using atoolbox called *Kaldi*, which must correspond as aclient created by an advanced programming language, with any home automation framework.

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

1. Introduction

The speech is the easiest way of transmission. Through it, we can voice out what we think. We can utilize it to showcase facts, thoughts, and feelings, wants to exchange, transfer, and ask for information. Today we don't simply utilize it to correspond with different humans, yet in addition with machines.

Speech recognition is the method that permits the examination of sounds got by a microphone to transcribe them into a progression of words that can be utilized by machines. Since its appearance in the 1950, automatic speech recognition has been frequently moved forward. Nowadays the applications of speech recognition are exceptionally differing and every framework has its own engineering and method of operation. The more extensive the field of application, the more prominent the recognition models must be "*keeping in mind the end goal to understand spontaneous talks and the assorted variety of the speakers*"[1].

In this paper, we will be interested in the automatic recognition of speech connected to home automation. In fact, it is the objective sought after by the savvy home which is a residence outfitted with PC technology to help its inhabitants in the different situations of the local life also as far as ease and safety are concerned. Automatic Speech Recognition could be vital benefaction to the perception of unusual circumstances, which is a vital piece of a home surveillance framework [2].This study is organized in the following parts: we begin by detailing the parts of a speech recognition framework. This is trailed by detailing parts of many speech recognition programs and testing their ability so as to pick the one, which will subsequently be amalgamated into a home automation framework.

2. Automatic Speech Recognition

A speech recognition framework is intended to connect a series of words with a series of acoustic observations. Accordingly, from the series of acoustic observations X, this framework looks for the series of words Ŵ which augments the probability P (W | X) that is the likelihood of emission of W knowing X [3].
The series of words Ŵ should then expand the 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 sorts of probabilistic modules are utilized to look for the most likely series of words: an acoustic model that gives the approximation of P (X | W), and a language model that gives the approximation of P (W). Fig. 1 demonstrates a general simplified diagram of the functioning of an automatic speech recognition system [4].

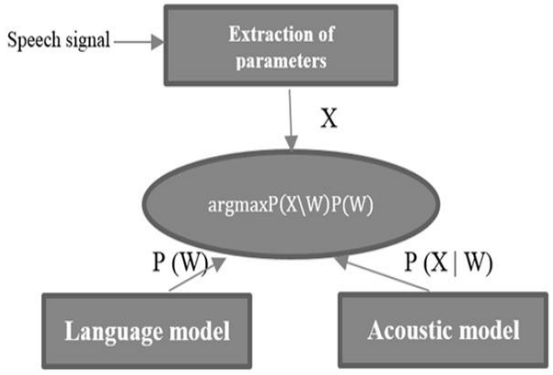


Fig. 1: Operation of a speech recognition system

As viewed in Fig. 1, the speech signal cannot be specifically changed into speculations of word sequences. The extraction of its parameter is an important advance since it must determine the necessary attributes of the signal. This extraction can be done using various methods, the most surely understood ones being parametric analysis, using the LPC (*Linear Predictive Coding*) strategy, the *cepstral* analysis, with for instance the MFCC *(Mel-scale Frequency Cepstral Coefficients*), or the PLP (*Perceptual Linear Prediction*) technique. These different strategies make it conceivable to extricate trademark coefficients for each frame. This extraction then makes it conceivable to obtain the sequence of acoustic observations X [5].

The acoustic model is a statistical model that gauges the likelihood that a phoneme has produced a specific series of acoustic parameters. A wide assortment of acoustic parameter series are watched for every phoneme because of all variations identified with speaker variety, age, gender, lingo, condition of well-being, psychological state. The most broadly utilized strategies for acoustic modeling are models in light of *hidden Markov models* (HMM) or *deep neuron networks* (DNN) [6].

Language models are forms that gauge the probabilities of the different word sequences P (W). These models are utilized to remember sequences of words from a literary corpus of learning. In the circumstance of speech recognition, language models serve to direct and constrain inquire about among different word hypotheses [7]. The most widely utilized language models are n-gram models.

So as to assess a few speech recognition frameworks, they ought to be looked at on a similar test information. Normally, these frameworks are assessed regarding word-error rates [8]. The WER considers the errors of:

- ✓ **Substitution**: Identified word set up of an expression of manual transcription.
- ✓ **Insertion**: Identified word placed in connection to the source transcription.
- ✓ **Deletion**: Word of forgotten source in the theory gave by the speech recognition framework.

The WER is expressed by the formula:

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

3. Automatic Speech Recognition Instrument

There are a few open-source programming for automatic speech recognition (ASR). Some noteworthy ones include HTK, Julius (both written in C), Sphinx-4 (written in Java), RWTH ASR toolbox and Kaldi (both written in C++) [9].

In the principal stage, we chose an ASR programming with the attributes adjusted to the constructionof our framework, or the slightest attractive to the best of every one of these needs. After we will portray the highlights of automatic speech recognition bolstered by this product.

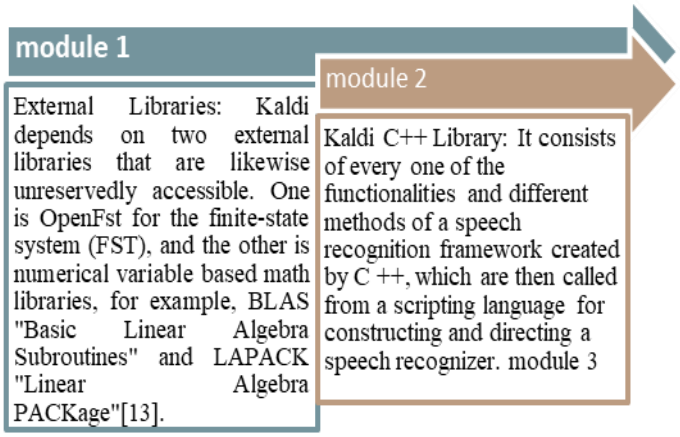
- **ASR Software: Kaldi**

To promote our investigation we pick the voice recognition program called *Kaldi*. It is an open-source toolbox for speech recognition written in C++ and licensed under the Apache License v2.0. [10]. This decision is spurred by:

- ✓ *Kaldi* have recent and adaptable code written in C++ that is very straightforward.
- ✓ *Open license*: The code is licensed under Apache v2.0, which is one of the minimum prohibitive licenses accessible.
- ✓ *Extensible design*: The calculations are produced in the most generic shape conceivable. This will enable us to effortlessly integrate our home automation.
- ✓ *Extensive linear polynomial math Support*: it include a matrix library that swathes ordinary procedures.
- ✓ *Finish formulas*: it make accessible finish formulas for building speech recognition frameworks that work from generally accessible databases.
- ✓ *Performance*: Kaldi exceeds the various recognition toolkits[11].

- **Review of Kaldi**

Kaldi is a speech recognition toolbox comprising of a library, command line programs and contents for acoustic modeling [12].The engineering of Kaldi, as portrayed in Figure 2, consists of the following modules:



with these two modules, we need to mention the following modules :

- ✓ *Kaldi* C++ Executables.
- ✓ Contents.

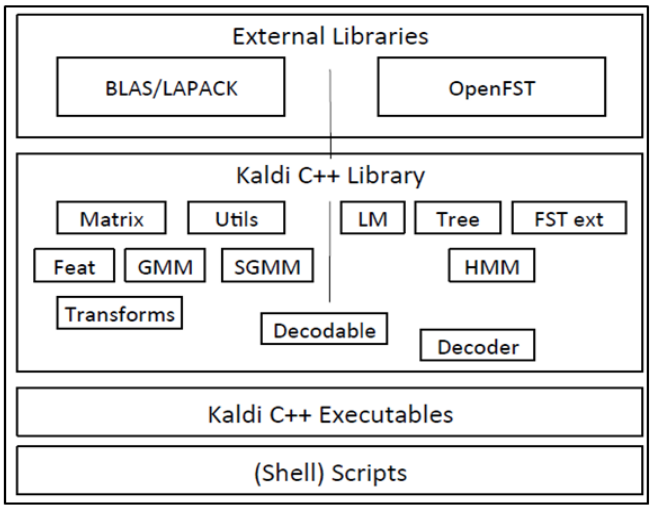


Fig. 2: Kaldi architecture

- **ASR characteristic in Kaldi**

As depicted in the second part of this paper, a speech recognition framework is made up of three modules: Parameter extraction, acoustic model, language model. The Kaldi library merges these three components and suggest the following codes:

- ✓ ***In case of the extraction of parameters:*** the Kaldi code goes for developing standard functions of MFCC and PLP, positioning sensible default esteems, yet giving individuals the opportunity to change the qualities.
- ✓ ***Concerning the acoustic model:***Kaldi bolsters normal acoustic models, for example, GMM, SGMM, HMM and DNN, it is additionally extensible to new sorts of models.

For languages models: Kaldi compatible with any language model that "FST". Therefore, it bolsters the most utilized n-gram model [14].

4. Kaldi for Domotics

The point of our work is to merge our selected speech recognition programKaldi, to a home automation framework. For this, we will initially propose a design of integration, which will be the focus of a few tests keeping in mind the end goal to qualify the performances of our model.

In an initial step, we propose a communication design between Kaldi and a home automation framework. Our proposition depended on a design in light of the OPC client/server technique (Figure 3) [15]. This decision is inspired by:

- ✓ The C ++ language supports OPC. In this way, Kaldi can be arranged as an OPC client [16].
- ✓ The client/server communication can be merged with the most home automation systems [17].

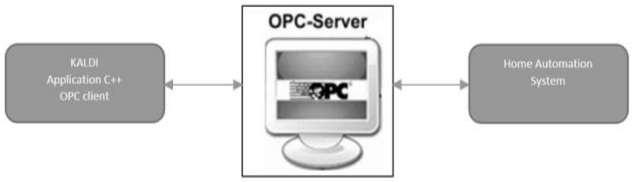


Fig. 3: Integration design

We will feature in this part the means created on *Kaldi* for the execution of the client/server OPCtransmission between our ASR program (called client) and home automation framework through its OPC server.The initial step is to connect to OPC Server;the following code will permit this connection:

```
HRESULT hr;
hr = CLSIDFromProgID(lpwSeverName, &clsid);
hr = CoCreateInstance(clsid, NULL, CLSCTX_ALL, IID_IUnknown, (LPVOID *)&pUnkn);
hr = pUnkn->QueryInterface(IID_IOPCServer, (LPVOID*)&m_pOpcServer);
hr=pUnkn->QueryInterface(IID_IOPCBrowseServerAddressSpace,(LPVOID*)&m_pOpcBrowse);
hr = m_pOpcServer->QueryInterface(IID_IConnectionPointContainer, (void**)&pCPC);
hr = pCPC->FindConnectionPoint(IID_IOPCShutdown, &m_pOPCConnPoint);
hr = m_pOPCConnPoint->Advise(m_pShutdownCP, &m_dwShutdownConnection);
```

The second phase is to make an OPC Group Object and Add Tags :

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

OPCITEMDEF* idef = new OPCITEMDEF[nPoints];
for(DWORD i=0; i<nPoints; i++)
{CItemObj* pItemObj = TempList->GetAt(pos);
CStringcsName = 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);
```

The final phase is to arrange read and write to and from an OPC server. For our situation, we will only arrange the write:

```
hr = m_pOPCGroup->QueryInterface(IID_IOPCSyncIO, (LPVOID*)&pOPCSync);
hr = pOPCSync->Write(1, phHandle, pItemValues, &pErrors);
```

5. Conclusion

After the exposition of the cutting edge automatic speech recognition systems,we portrayed the design of *Kaldi*, a free and open-source speech recognition toolbox. It bolsters an extensive variety of strategies for extracting parameters, acoustic models and language models.

We were likewise ready to arrange *Kaldi* as an OPC client to have the capacity to merge it into a home automation framework through its OPC server. Further work will focus on executing this integration keeping in mind the end goal to test the toughness of our framework.

Acknowledgement

This work has been supported by the grants of National Centerfor Scientific and Technical Research (CNRST- Morocco): No. 757UIT for the fourth author. The authors would like to thank all committee of the International Conference of technology, Innovation and Information systems (CITISI/2), for their organization and supports.

References

[1] Allauzen et J.-L. Gauvain, «Construction automatique du vocabulaire d'un système de transcription,» Journ{\'e}es d'Etude sur la Parole, 2004.
[2] M. Vacher, «Analyse sonore et multimodale dans le domaine de l'assistance à domicile,» 2011.
[3] R. Dufour, «Transcription automatique de la parole spontanée,» 2010.
[4] M. Bouallegue, «L'analyse factorielle pour la modélisation acoustique des systèmes de reconnaissance de la parole,» 2013.
[5] Z. Le-Qing, «Insect sound recognition based on mfcc and pnn,» chez Multimedia and Signal Processing (CMSP), 2011 International Conference on, 2011.
[6] F. Bougares, «Attelage de systèmes de transcription automatique de la parole,» 2012.
[7] P. Karanasou, «Phonemic variability and confusability in pronunciation modeling for automatic speech recognition,» 2013.
[8] N.-T. Le, C. Servan, B. Lecouteux et L. Besacier, «Better Evaluation of ASR in Speech Translation Context Using Word Embeddings,» chez Interspeech 2016, 2016.
[9] F. Aman, M. Vacher, F. Portet, W. Duclot et B. Lecouteux, «CirdoX: an On/Off-line Multisource Speech and Sound Analysis Software,» chez Language Resources and Evaluation Conference, 2016.
[10] S. Madikeri, S. Dey, P. Motlicek et M. Ferras, «Implementation of the standard i-vector system for the kaldi speech recognition toolkit,» 2016.

[11] Gaida, P. Lange, R. Petrick, P. Proba, A. Malatawy et D. Suendermann-Oeft, «Comparing open-source speech recognition toolkits,» Tech. Rep., DHBW Stuttgart, 2014.
[12] Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz et others, «The Kaldi speech recognition toolkit,» chez IEEE 2011 workshop on automatic speech recognition and understanding, 2011.
[13] Povey, M. Hannemann, G. Boulianne, L. Burget, A. Ghoshal, M. Janda, M. Karafi{\'a}t, S. Kombrink, P. Motl{\'i}{\v{c}}ek, Y. Qian et others, «Generating exact lattices in the WFST framework,» chez Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on, 2012.
[14] C. Allauzen, M. Riley, J. Schalkwyk, W. Skut et M. Mohri, «OpenFst: A general and efficient weighted finite-state transducer library,» chez International Conference on Implementation and Application of Automata, 2007.
[15] O. Passalacqua, E. Benoit, M.-P. Huget et P. Moreaux, «Integrating OPC Data into GSN Infrastructures,» arXiv preprint arXiv:0808.0055, 2008.
[16] H. N. Li Zheng, «OPC (OLE for process control) specification and its developments,» 2002.
[17] Topalis, G. Orphanos, S. Koubias et G. Papadopoulos, «A generic network management architecture targeted to support home automation networks and home internet connectivity,» chez Consumer Electronics, 1999. ICCE. International Conference on, 1999.