An Automatic Speech Recognition Solution For Home Automation

M.ABOULKHIR [a], S.BOUREKKADI[b], S.KHOULJI [a] ,A.BABOUNIA, M.L.KERKEB[d]

[a,] ERISI, Abdelmalek Essaadi University, National School of Applied Sciences, 93000,Tetouan,Morocco

***Abstract*—. This research work analyses the various Automatic Speech Recognition frameworks linked to home automation. In order to achieve this, a detailed investigation of the engineering speech recognition framework was carried out. The aim is select a speech recognition program capable of working in remote speech regions as well as in over populated regions. An ASR toolbox known as Kaldi is the ASR toolbox used, which is corresponds with the OPC client developed in C++ and implemented in home automated framework. The OPC acts as a sever.**

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

I. Introduction

Speech happens to be the easiest means of transmission. Speech enables one to communicate, showcase feelings, facts, thoughts, and lots more. Speech is currently not only used as a means of communication, by humans but also as a means of transferring and transmitting information to machines.

) Speech recognition can be defined as the technique used to transcribe the sounds obtained from a microphone into progressive words that can be translated by the machine. Since its inception in 1950, automatic speech recognition has been consistently advanced with its applications exceptionally differing with regards to its mode of operation and engineering design. The more comprehensive the field of application, the more distinguished and well-known the recognition model is expected to be [1].

In this research work we will focus in the automated recognition of speech linked to home automation. This is the sole objective of this research work which happens to be the sole objective of savvy home when seeking for the best technique to assist its inhabitants in their everyday life. With the help of the PC technology, it will assist habitants with the day-to-day local activities. Automated Speech Recognition could be described as an invaluable benefaction to the recognition of unusual circumstances, which happens to be an imperative piece of all home surveillance architecture [2].

In this research work, we start by detailing the various parts of a speech recognition architecture. This is done by analyzing various parts of speech recognition schemes and testing their capability in order to pick the ideal one that would be incorporated into a home automated framework.

II. Automatic Speech Recognition

A speech recognition scheme is meant to connect series of words with that of observations. Furthermore, from the series of sound observations X, this scheme tends to resemble series of words depicted with W, which results to the probability P (W | X). This probability emits W with respect to X [3]. The series associated with W, thus expands as follows:

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

Applying the Bayes rule, we obtain the formula

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

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

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

Two kinds of probabilistic modules are used to obtain the right series of words know as an acoustic model that results to the approximation of P (X | W), as well as a language model that results to a language model which leads to a language model of P (W). In Fig.1, an equation which simplifies the functioning of an automatic speech recognition system is demonstrated and simplified [4].

As displayed in Fig 1, it is impossible to change the speech signal into word sequences. This is because the extraction of its numerous parameters which is eliminated is extremely important to its advancement. The extraction can however, be done using various techniques with the Linear Predictive Coding (LPC) technique, Mel-scale Frequency Cepstral Coefficients (MFCC) technique, and the Perceptual Linear Prediction technique. These various techniques make it pretty easy to extricate trademark coefficients for each frame thereby making it possible to obtain the various acoustic observations X [5].

The acoustic model is described as a statistical model that measures the likelihood that a particular technique has the capability or is capable of producing a specific series of parameters. Owing to the numerous numbers of variations

IJSER

identified which is influenced by a couple of factors such as age, lingo, gender, psychological state and lots more the Markov models and the Deep Neuron Networks (DNN) are the most widely utilized acoustic strategies [6].
Language models are bench marks that gauge the probabilities of various word sequences P (W). These models are used to remember sequences of words from a corpus of learning. Language models directs and constrains inquires on the various word hypothesis of speech recognition [7].
In order to assess a couple of speech recognition schemes, they are meant to be assessed on a similar test information framework. Normally, these schemes are evaluated with reference to word-error rates [8]. The following errors are considered by the WER:

- ✓ Substitution: Recognize words that have been set up of an expression of manual transcription.
- ✓ Insertion: Recognize words that are placed in connection to the source transcription.
- ✓ Deletion: Word of omitted source in the theory issued by the speech recognition scheme.

The WER is signified by the formula below:

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

III. AUTOMATIC SPEECH RECOGNITION INSTRUMENT

There are a couple of open-source programming used for automatic speech recognition (ASR). Some well-known ones include Julius and HTK which are both written in C, Sphinx-4, RWT ASR toolbox and Kaldi which is written in C++ [9]. The ASR programming language is used at the principal stage. Its attributes are adjusted to fit into our framework. After that we will proceed to depict the highlights of automatic speech recognition expedited by this product.

ASR Software: Kaldi

To advance our investigation we made use of Kaldi, an open-source voice recognition program written in C++ and licensed under the Apache License [10]. This incited by:

- ✓ Kaldi is written in C++ which happens to be an adaptable and straightforward programming code.
- ✓ It is an open license licensed under Apache v2.0, which happens to be one of the least prohibitive licenses that is accessible.
- ✓ Extensible design: The calculations are carried out in the most reasonable and understandable wayThis will enable us to easily consolidate our home automation.
- ✓ It includes a matrix library that extensively support linear polynomial and which swathes simple procedures.
- ✓ Finish formulas: It provides formulas for building speech recognition frameworks that works on all databases.
- ✓ Performance: Kaldi surpasses the numerous recognition toolkits [11].

Review of Kaldi

Kaldi can be defined as a speech recognition toolbox that consist of a library, command line programs as well as contents for sound modeling [12]. Kaldi engineering concept consists of the following modules:

- ✓ External Libraries: Kaldi is dependent on two external universally accepted libraries. The first is known as OpenFst which is used for the finite-state system (FST), with the other being a numerical variable based math libraries, for instance a, BLAS "Basic Linear Algebra Subroutines" and LAPACK "Linear Algebra PACKage"[13].
- ✓ Kaldi C++ Library: This library consists of individual functionalities as well as various methods of a speech recognition framework designed and created by C ++.
- ✓ Kaldi C++ Executables
- ✓ Contents

Figure 1 : – Kaldi architecture

ASR characteristic in Kaldi

As detailed in the second section of this research paper, a speech recognition framework consists of three modules: The Acoustic model, Parameter model, and the Language model. The Kaldi library conglomerates with these above listed components to produce the following codes:

IJSER

- ✓ With regard to the extraction parameters, the Kaldi code is used for developing standard functions of PLP and MFCC. It positions sensible default esteems while giving individuals the fortuity to change qualities.
- ✓ With reference to the acoustic model the Kaldi ameliorates normal acoustic models for instance GMM, SGMM, HMM and DNN.
- ✓ Regarding the language models, Kaldi permits us to bestow any language model that can easily be portrayed as a FST thereby ameliorating the most utilized n-gram model [14].

IV. KALDI FOR DOMOTICS

The sole aim of our research work is to consolidate Kaldi (a speech recognition programme) into a home automation framework. As a result of this, we propose a design integration, which will be the focus of a couple of few tests keeping in mind the end goal.

- ✓ In an introductory step, we recommend a communication design between a home automated framework and Kaldi. The OPC is the dependent server used. The decision is influenced by the fact that the: C ++ language supports OPC, thereby making it possible for Kaldi to be arranged as an OPC client [16].
- ✓ The client/server communication can be consolidatedwith the most home automation systems[17].

The second phase involves making an OPC Group Object followed by the addition of Tags

The final phase involves the arrangement of read and write to and from an OPC server.

V. CONCLUSION

After proper detailed description of the automatic speech recognition systems, we proceeded to the design and development of a free open source speech recognition toolbox known as Kaldi. A toolbox which aims at ameliorating a wide variety of strategies for extracting acoustic models, parameters and language models. We are also ever ready to arrange the Kaldi as an OPC client in such a way that it has the capacity to merge a home automated framework using its OPC server. Further extensive research will aim at executing this integration while keeping in mind the result in order to test the adamancy of our proposed scheme.

```
// 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);
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);
```

Am actually still working on the remaining two articles, need some extra days like three more days. Hope that would be cool with you.

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

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

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

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

[7] 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.

[8] 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.

[9] 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

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

[11] Zheng, L., & Nakagawa, H. (2002, August). OPC (OLE for process control) specification and its developments. In SICE 2002.

IJSER

ISSN 2229-5518
Proceedings of the 41st SICE Annual Conference (Vol. 2, pp. 917-920). IEEE.

[12] 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.

IJSER