Designing the Workflow of a Language Interpretation Device Using Artificial Intelligence

Jabed Hossain[1], Md. Aminur Rashid [2]

[1](Department of Computer Science and Information Technology, American International University-Bangladesh, Bangladesh)
[2](Department of Computer Science and Information Technology, American International University-Bangladesh, Bangladesh)

Abstract : *Among all the obstacles there existed and currently existing between mankind from very early age of human civilization to the age of modern globalization, language difference between humans has always been a crucial problem being a spine in the throat. This paper emphasizes on language interpretation device and its application in the modern world to alleviate the difference between humans and the discrimination between tongues within continents. The paper deals with a concept of HCI (Human Computer Interaction) in developing a language interpretation device which can automatically interpret multiple languages into a target language in a dynamic environment within a specific time. The concept is developed with a sense of mobility into action with an efficient method of consuming voice as input and developing a processed meaningful voice as output. The paper also demonstrates the work flow of the systems those are implemented in this device. The paper also describes the core system architecture and the function of component in the system. The effectiveness of the presented approach also is shown with some future works.*

Keywords: *Language interpretation, HCI (Human Computer Interaction), SRS (Speech Recognition System), mobility, language cartridge.*

I. INTRODUCTION

The distance between the continents, sub-continents and countries of the world is shrinked due to globalization. Travel between countries is strikingly increased for pleasure and also for business perspective. One of the main problems associated with travelling is language barrier. The numbers of pairs of languages and documents, between which translations must be pertained, is increasing day by day. There is not enough money to invest on the profession of a translator with an efficient and optimistic output.

If language cannot be understood, the distance between people gets large; if the distance is too large then the people will be unsocial which is contradictory with the nature of humans. In recent years language translation and interpreting industry has been described using terms of Globalization, Internationalization, Localization, and Translation [1, 2, 4, 5].

The phrase 'going global' is no longer used for getting attention, but it has now become a fact in day to day life. Every moment we live we run towards globalization and play our part in the world economy, for instance, we drink coffee which is imported from some South American countries, the vehicles that we use are from Germany and the oil or gas is imported from Arab countries. Within all these importing and exporting transaction, interpreting devices could make substantial impact through interpreting transfer manuals, documents, agreements and conversations from source to target languages.

The invention of such devices which can interpret languages is yet to be revealed because there are not enough evidences of successful utilization of these kinds of devices. Martin *et* al. [5] verified the fact that over the past twenty years scientists have been trying to co-ordinate computer science with computational linguistics to come up with some instructive results. This led the invention with no such result which can be applied to clone some other devices too but scientists found some result under very special circumstances which is not a very effective success because it is contradictory to the real environment where these devices are supposed to work[5]. Omar *et* al. [3] proved the necessity of implementation of ROILA (Robot Interaction Language) with respect to communicate with different devices which is in this case a device which will be operated as a computer or CPU based robot. Omar *et* al. [3] also described about HRI (Human Robot Interaction), Speech Recognition and also HCI (Human Computer Interaction) which are generally used as components of artificial languages to interact with robots.

In this paper at the "Proposed approach" section an algorithm is proposed using a speech recognition system, a target language generator, a word match score generator and a translation match score generator, which will work to translate a target language from a source language. The device can posses a database, which will be trained to enrich itself with phrases, words and their definition to work in real environment. Also a voice recognition module can be included which will recognize the voice and accent of a particular speaker and will

fetch translated language from the system [2]. The implementation of SRS (Speech Recognition System) within the CPU (Central Processing Unit) is also described in the training stage of the proposed section. Implementation of HRI using ROILA also described in the future work section. The paper also describes a method by which a language interpreter will train itself in the working stage to evolve with dynamic environment, in the proposed approach.

II. RELATED RESEARCH

The countries of the world inhabited by different citizens who speak different kind of language and idioms, do generally make sense to a very few people all over the world. The difficulties associated with reading distinct languages and interpreting those speaker's accents and whose language the other speaker cannot interpret, creates a language difference between peoples.

Unfortunately, while some illustrious success has been achieved in this orbit, little attention has been given to the automated devices those can be used as a language interpreter. Traditional voice language translator generally includes a display system, language cartridge, a speaker, a microphone, a voice synthesizer, a voice recognition module, and a programmed CPU [1, 2]. Developed to act as a translator the device is trained to the voice of a voice actor [2]. In recent invention the translator or interpreter can possess a pre-programmed voice in its RAM which will be employed by the CPU [4].

In recent inventions scientists are more interested to built even smarter device or robot to be used as an interpreter. Within the context of Multimodal Human Computer Interaction, Speech Interaction is one of the interaction modalities. Technologists are training devices with HRI (Human Robot Interaction) for recognizing the SRS (Speech Recognition System) [3].

In this paper, a model of a translator is proposed which is physically similar to the traditional model structure of an interpreter but the workflow of the system will be somewhat different than the traditional ones [1, 2]. The difference will be in this paper an Artificial Intelligence system is proposed to be implemented with the HCI implementation for speech recognition in the device. The design which is unveiled in this paper will also focus on the smartness of the automated interpreter. It is proven on researches that implementation of HCI improves performance on the training period of an interpreting device which was rare in the traditional devices [4, 5, 6].

III. PROPOSED APPROACH

According to the concepts of previous acquisition regarding interpretation device, this paper proposes to approach the problem with two different stages of implementation which are training stage (pre-working, initial stage) and working stage. Most of the concepts of system architecture of the interpretation device are constructed from previously existing acquisition regarding language interpretation, ROILA(Robot Interaction Language), speech recognition as well as speech translation task with limited changes as possible [2,3,5].

3.1 Device Structure

Opining of such a device, the first drawback that generally pop up into the scenario regarding an interpretation device is the deficiency of mobility. As, a language interpretation device is to be developed for winnowing out of interaction distance between inhabitants all over world, it has to be gifted with an special adjective "Great Mobility". Because without containing mobility, the interpretation device will not be efficient enough to produce error free output in a real environment. Stephen *et* al. [2] developed an elementary design with the implementing the sense of mobility. In this paper an approach has been taken to explicate the theory of design with limited changes and with different system architecture [2]. The device will consist of a programmed Central Processing Unit (CPU), which will control the advancement of processes which will be run by applying distinct processing algorithms. Central Processing Unit (CPU) will also comprise an efficient Random Access Memory (RAM) which will enable the device to work faster by producing error free output. There will also be a voice synthesizer which will be responsible for imitating variety of voices. The device will contain a display screen which will be used as an active user interface run by the CPU. The device will also feature a speaker and a microphone will also be installed in the hardware section.

3.2 Training Stage

The training stage is considered to be the pre-working stage too. In this stage the device will make itself worthy of becoming an efficiently utilitarian object which is beneficial for the well being of human communication above all barriers. From the structural design provided earlier we have decided to use storage device with maximum data storing capacity. The training stage generally refers to training the CPU. The storage chip which is to be installed in the CPU, will consist a large number of data regarding interpreting languages. In the training stage the device will follow efficient working algorithm which will be programmed into the Central Processing Unit to control the work flow.
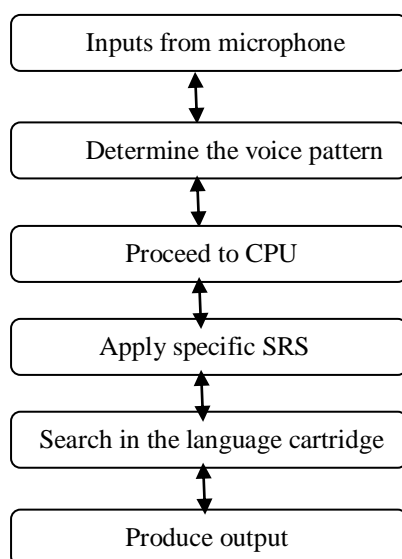
Fig: 1. *Work flow structure of a language interpretation device.*

At first, the device will listen to the human speaker through a microphone. During the delivery of the speech of a speaker the device will determine the voice pattern of the speaker through voice recognition system [2, 3]. A device's recognition of speech generally refers to enabling itself to "recognize" what was spoken. Incoming audio vibration is matched against stored decoration of speech which represents various sounds in the language such as words, phonemes or other similar sound units and sound patterns might also represent linguistic constraints. The voice recognition system will implement output considering the vibration of the tone of the human speaker. Considering the emotion of the human speaker, there might be several types of tones which will be encountered by the interpretation device such as anger mood, interrogative mood, happy or funny mood, sad mood etc. It will most likely be an epochal decision making process from the point of view of the interpretation device as it will have to perform in the real scenario within different constraints. So every emotion behaved by the human speaker will have to get back to him accordingly so that communication hits the tempo without any vacillation.
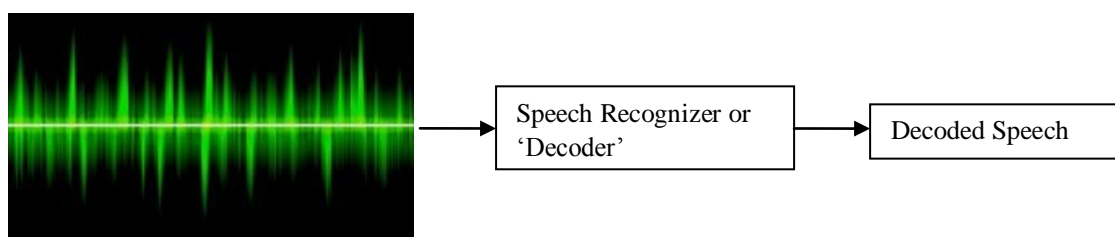

Fig: 2. General scenario of a speech decoder**.**

After determining the voice pattern Central Processing Unit will work to employ a speech recognition system. There are many speech recognition systems available which are well implemented regarding interpretation of language. Murveit *et* al. used a version of SRI's DECIPHER speaker independent continuous speech recognition system to a good upshot [6]. Similar kind of speech recognition system is used to train the Central Processing Unit. Through a speech recognition system the CPU will be trained with severe amount of sentences and vocabularies. SRI's DECIPHER generally uses FFT-based Mel-cepstra font end.

If the CPU successfully recognizes the speech then it will compare that speech with the language speech pattern stored within the language cartridge [2]. This process will be quite pugnaciously engrossing as in this step all the step result will be summarized to propel a competent output. The speeches which are collected from the user will be examined to assay a match of speech voice pattern with the stored patterns in the cartridge to fetch efficient interpretation. As for an example, if the user throws an interrogative sentence, the CPU will first concentrate on the voice pattern and then will work on speech recognition system. After these processes are completed the CPU will watch for the match in the language cartridge. This process is only done if the previous two processes are completed and the result it produces comes from summarizing all the previous outputs. This step is mainly because for optimization purpose. It will reduce the rules to propel a competent result.

3.2.1 Algorithm:

According to the description earlier, an algorithm can be proposed to affirm the described method and its effectiveness. Followed by the work flow structure described in 'Fig. 1', after getting a voice speech as input, from the environment the device can store it in a voice array.

```
Begin:
 language: = array (English, Bangla, French…); //Language stored in an sound array.
speaker: = array (female, male); //Stored voice of male and female
 input: = array (); //Stores the speech from a speaker.
 *Apply SRI'S DECIPHER speech recognition system
function_check_speaker (speaker [], input []) // checks whether the speaker is male or female.
{
// Checks whether the input data are from a Male or a Female by simulating the sound vibration or sound
wave
  if Male
    return  m;
  else
    return f;
}


function_check_langauge (language [], input []) // checks the language of the speaker.
{
// Checks whether the input data in the input array are similar to any data in the language array.
  if (input [] is subset of language [])
          return language []; //select the language which is similar to input []
  else
      //Insert input ["input language"] array information into a new index of the language []
      return language []; // updated language [] array.
 }
function output (function_check_speaker (), function_check_language())
{
  if (// if the function, function_check_language () returns language ["input language"] array)
  {
   Select the target language;
   if ("Selected language (Bangla, English…)")
   {
    //Apply target hypothesis generator;
    //Apply word match score generator from the language cartridge;
    //Apply translation match score generator from the language cartridge;
    output = array (); //stores the translated output;
   }
   else {
       break;
       }
  return output ["translated language"];// returns the translated language;
  }
 else
 {
 //target language does not exist;
 }
}
End;
```

3.3 Working Stage

The interpreting device which has been designed will be capable of learning from the working environment and enrich its language cartridge effectively with consistency. As the device will be used for interpreting different language it is very obvious that it will confront some problems translating all the sentences, words as well as phrases. It will possess a recorder which will record all the conversation of a day or

of a month. Then it can easily be derived which elements of language were missing in the cartridge for interpreting different language and instantly can be upgraded to solve those problems. Thus the device will upgrade itself while working dynamically.

IV. FUTURE WORKS

As it is discussed earlier, now a days; HRI system (Human Robot Interaction) has been fueled to top to interact with machines by humans [3]. Human robot interaction now reaching in such a level that people can now interact with machines just like they interact with other people. Several HRI systems are included in the CPU of modern devices with minimum uncertainty and maximum optimization now a day, so that the machines are livelier than ever. A HRI system can be implemented to the interpretation device as it is likely to be developed for wider use rather than only in corporate use only in the future. Robot Interaction Language (ROILA) is another measure which can be brought up to the scenario for making the interpreting device more user friendly.

V. CONCLUSION

Interpreting devices now a days, generally interpret after a speaker speaks but not simultaneously. What if a robot can talk simultaneously with a person or what if a robot can answer questions using artificial intelligence just like a human does? The difference between humans will be alleviated and discrimination between languages will be removed for good.

In this paper the possibility of producing such a device which can learn through effective procedures from real environment to perform simultaneously with minimum time using some effective steps described in 'Fig: 1', has been discussed. The time needed to respond against human speech by interpreting the speech needs to be lessened to achieve the goal of interpreting simultaneously rather than continuously. It will be really feasible and someday it will be possible.

Acknowledgements

REFERENCES

[1] J. M. Lande. Interpreting Device. *United States Patent Office, Serial Number. 506,603, 3 Claims (Cl. 35-2)*, 1943.
[2] Stephen A. Rondel, Redmond and Joel R. Carter, Mukileto. Voice Language Translator. *United States Patent Office, Application Number. 306,001,*1989.
[3] Omar Mubin, Cristoph Bartneck and Loe Feijs .*Towards the design and evaluation of ROILA: a speech recognition friendly artificial language. Advances in Neural Processing.* Pages: 250-256, 2010.
[4] Xi Shi and Yangsheng Xu. A Wearable Translation Robot. *In Proceedings of the 2005 IEEE International   Conference on Robotics and Automation,* 2005.
[5] Martin Kay. The Proper Place of Men and Machines in Language Translations. *Machine   Translation, Volume: 12, Issue: CSL-80-11, Publisher: Springer,* Pages: 3-23, 1997.
[6] Hy Murveit, John Butzberger, and Mitch Weintraub. Speech Recognition in SRI's Resource Management and ATIS Systems. *Speech and Natural Language, Proceedings of a Workshop, HLT,* 1991.