

A Gesture Audio-Video Conferencing Application for the ease of Communication between Normal Person And Deaf and Dumb Person

Nikhil T. Tarte

Dept. of Computer Engineering,
Sinhgad College of Engineering,
Vadgaon (BK), Pune – 41

Sneha N. Bhadane

Dept. of Computer Engineering,
Sinhgad College of Engineering,
Vadgaon (BK), Pune - 41,

Purva S. Kulkarni

Dept. of Computer Engineering,
Sinhgad College of Engineering,
Vadgaon (BK), Pune – 41

Abstract - In today's Era, a person with Disability is not able to stand in a race with a normal person. Considering a person with lack of hearing and speaking ability deprives this race because of lack of proper communication between them. This paper gain is to present at the glance to the application proposed, to overcome the problem for such people by an application viz. Development of a Gesture Audio – Video Conferencing Application, which enhances the communication between a Deaf and Dumb Person and a Normal Person without any requirement of a special Sign Language Interpreter. This application is responsible for translation of Naturally Spoken English Sentences in form of Speech to Visual Sign Language Gestures and from Gesture to audible Voice in form of Natural English words. It consists of a major modules like to realize this an explanation of internal working of such an application constituting Gesture Template Matching and Natural Language Processing as Most Important Module is presented in paper. In a narrow spectrum it acts as a Language Interpreter and provide a convenient way for Communication, and responsible for a step ahead in the direction of a new Era, where disability is not a barrier.

Index Terms— Gesture Recognition, English Grammar, Indian Sign Language (ISL), Voice Synthesis, Voice recognition, Natural Language Processing.

I. INTRODUCTION

India constitutes 2.4 million of Deaf and Dumb population, which holds the world's 20% of the Deaf and Dumb Population. These people lacks the amenities which a normal person should own. The big reason behind this is lack of communication as deaf people are unable to listen and dumb people are unable to speak. Fig. 1 shows a survey analysis. This decreasing ratio of Literate and Employed Deaf and Dumb population is a result of the physical disability of hearing for deaf people and disability of speaking for dumb people so it yields to lack of communication between normal person and Deaf and Dumb Person. It actually becomes the same problem of two persons which knows two different language, no one of them knows any common language so its becomes a problem to talk with each other and so they requires a translator physically which may not be always convenient to arrange and this same kind of problem occurs in

between the Normal Person and the Deaf person or the Normal Person and the Dumb person. To overcome this problem, we introduce a unique application. Our application model is a desirable Interpreter which translates (1). Natural English Sentences as, an Audio Input by Normal Person to Indian Sign Language i.e. Meaningful words in

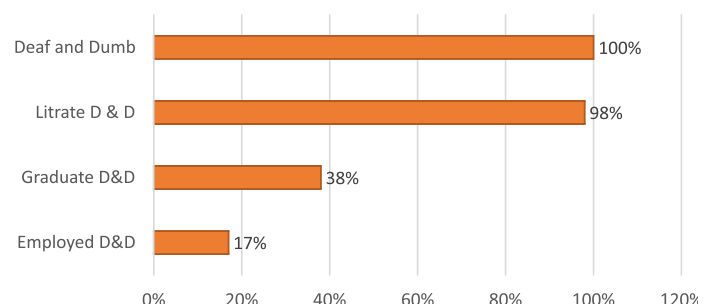


Fig. 1: Deaf and Dumb Work Survey

the form of Animation or Video for Deaf Person and (2) Sign Language, in form of Gesture by a Dumb Person to Synthesized English Words which have a corresponding meaning in Sign Language which interprets a particular thing, as an Audio Output for Normal Person. This will help Normal and Deaf and dumb communities by removing the communication gap between them.

To implement such a versatile model to a realistic application requires a series of Modules, viz.(1) Gesture Recognition for Sign Language Input, (2) Natural Language Processing (3) Voice Synthesis and Recognition for Input and Output of Audio, (4) A Database Organization and further more.

Before moving further, we must know some basics. Firstly, what is a Language, its Syntax and grammar in which an information is carried out English.

Understanding of Subject and Object in a Sentence is itself a Major Challenge.

The exact goal and its implementation can now be explained as below in Fig. 2, this will have a different working

for both users; i.e., Normal Person and Deaf and Dumb Person.

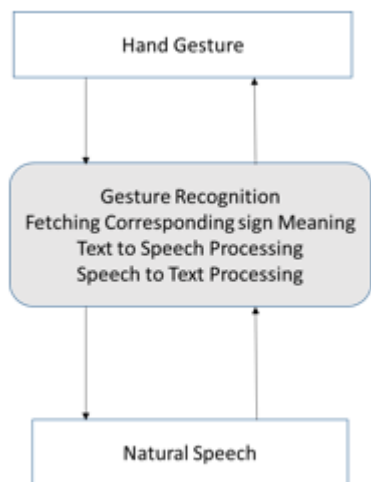


Fig. 2: The Working of the Application

The Process Implementation consists a loads of Challenges as (1) there are around 2000+ gestures standardized by Indian Sign Language Community and most of the gestures dynamic and some of the gestures are in static format, example there are 26 alphabets are in English language and 10 numbers, each alphabet character and a number has an particular sign in sign language and they are in static format i.e. in one image those gestures can be shown but there are many other gestures which are in dynamic format, example –he, she, about, dress, apple, Diwali and may more which are very promising and challenging to process. (2) Complex sentences have very difficult format. So deformation of complex sentence and converting them into sign language is one of a challenging task to do. (3) There are some verbs which have same gesture for its present tense, past tense, future tense and some verbs have different gestures for its past tense, present tense, future tense. E.g. Go. (3) Database Organization, as it is known that there are 2000+ gestures are standardized by Indian Sign Language Community it is greatly time consuming to find out the synonyms for each gesture so that it should not happen that for two words having same meaning, but one word has gesture in database and second word don't have a gesture in database. (4) Handling all types of tenses is very difficult process. (5)

The strategy of development of this is explained below Module wise.

II. LITERATURE SURVEY

For the Deaf and Dumb people, a lots of research is ongoing for gesture recognition..A method had been developed for recognizing isolated Arabic sign language gestures in a user independent mode. In this method the signers wore gloves to simplify the process of segmenting out the hands of the signer via colour segmentation. Many researchers utilized special

devices to recognize the Sign Language. Hand shape and motion are extracted easily and accurately using these devices. However, devices were expensive and, crucially, they reduced the naturalness of Sign Language communication. Our approach is without using any devices for recognition. Instead we are using a depth gesture recognition camera which is not much costly compared to other devices.If we refer the previous work for Gesture Recognition and Language handling, it will surely give us direction for the development of this application. But to bring out something very helping and creative software there is always need of innovative ideas. This is how we are trying to implement our own algorithms for creating this software.

III. IMPLEMENTATION

The Application has the two end users; i.e., Deaf and Dumb End, illustrated in Fig. 3 and Normal User End illustrated in Fig. 4.

The work flow in Fig. 3 states the working implementation of how this is going to get carried out. The first step is Input of Gestures. These gestures are taken as a 32- bit blob depth Image by the Intel Depth Camera and processed by Intel Perceptual Computing SDK to Extract the Coordinates of hand Movements. It gives 3D co-ordinates i.e. X, Y, Z co-ordinates. As mentioned in above introduction we have most of the gestures in dynamic nature, and some of are in static nature, so for static gestures only 2D co-ordinates of an image is required but for dynamic gesture recognition the centric parameter i.e. Z co-ordinate is required to analyze the gesture. These coordinates are processed by tracking their Pose and Gesture from Fingers to Hands. Later, Translation of ISL to English is carried by NLP Engine with the help of Learned Rules and Database of words

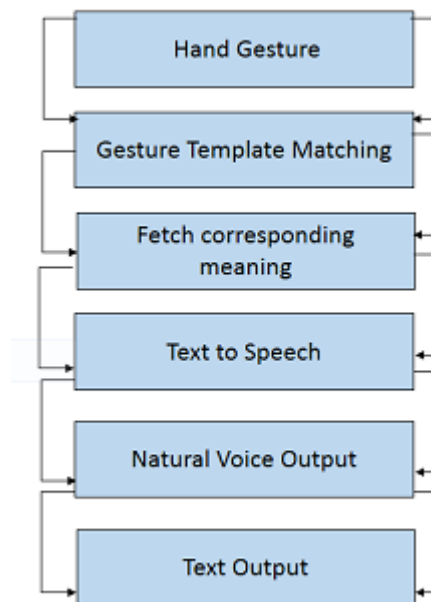


Fig. 3: Dumb User End to Normal User End

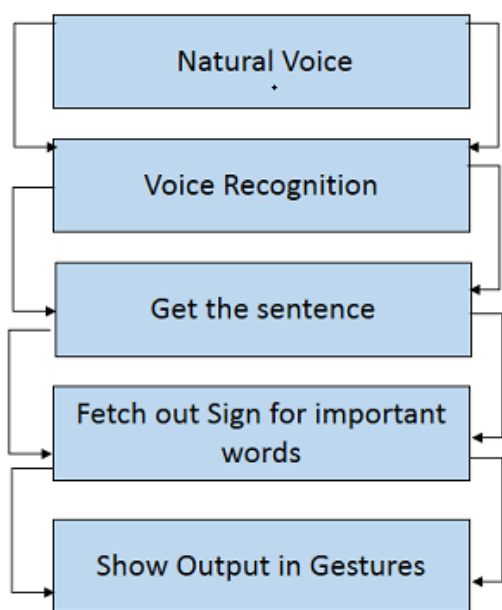


Fig. 4: Normal or Dumb Person to Deaf and Dumb Person

. These Sentences are then spoken by a Computer using Voice Synthesis. On the other hand, Fig. 4 states the Translation of natural sentences into gestures. For this the first step needed to be carried out is Voice Recognition. The voice is required to be recorded to its very first stage. This voice is then sent over the network to Google Voice Recognition System. This will get processed and then a Text is received. The main Reason of using Google Voice recognition API is it has a very High Precession for Indian English Pronunciation. At the last stage these sentences required to be fetched from database, in form of animation or Video.

The implementation work of this application owns a self-great reliability and efficiency, by its Designing Phase.

- Machine Learning – The Natural Language and Gesture Recognition Algorithms are designed in a New Way which Learn by itself.
- Voice Recognition – We have used Google API to perform the same keeping in mind the Indian English Accent and Accuracy of Google Voice Recognition Service.
- GUI – Our GUI is designed in Windows Presentation Foundation (WPF) Client for a very high flexibility and Ease to the User.
- Documentation – The Documentation of the Whole Application is done properly and responsible for better Usability.
- Chat Like Interface – The Application is made like a Conferencing application illustrating make best usability of such a Concept.

Each of its module consists its own heavy importance and a unique style of working, which is stated in following sections.

IV. NATURAL LANGUAGE PROCESSING

At both the ends our project takes natural sentence as input or as output so it becomes important part to handle the natural sentences and language. Here we are using English Language as a natural language and its syntax. Natural Language Processing is the heart of our Application and is the most important and challenging module, which is required to be performed.

V. GESTURE RECOGNITION

On the basis of the movement of the hand in the 3-D environment. Each movement or a static instance gives a set of hand and Fingers features.

After recognition of the coordinates the track of the movement is carried out for dynamic gestures and an Instance is recorded for a static gesture.

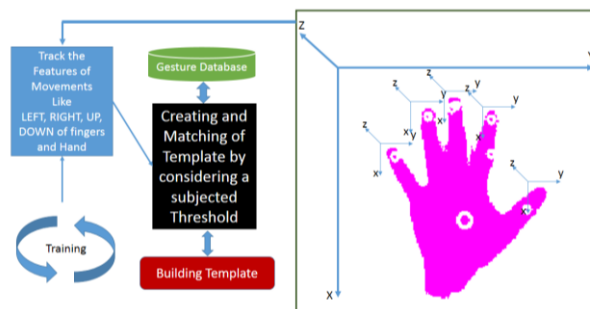


Fig. 6: Gesture Recognition

The features will be as moving of left hand in right direction and thumb up. These sets of features create a meaningful ISL word. And the sets of such meaningful words creates a meaningful ISL Sentence.

Fig. 6 states the process of gesture recognition for training and Identification of gestures. As the Hand and fingers motion is tracked it matches with its earlier instance and results an Identified Gesture. It does a self-training of the gesture to improve its efficiency at later states. This is as when a gesture is performed and if it provide an incorrect output it track user corrected result and improve its effect.

VI. VOICE SYNTHESIS AND RECOGNITION

The Voice synthesis is the Text to Speech Conversion. This is to provide a complete and proper conferencing touch. At the Normal and Dumb person end it will effective. The Voice synthesis is carried out by Intel Perceptual SDK Voice synthesis module.

At the other end Voice recognition is conversion of Speech to Text. This is required at the Normal and only Deaf end for recognition of sentences. The Speech of Indian English is having a Different and so we had performed the Speech to Text conversion by Google Speech Recognition API. This takes a recorded speech sentence and provide the text form of it for processing. This both system makes the application a complete conferencing application.

VII. APPLICATIONS

This application itself is responsible for a machine Sign Language Interpreter which can be available any time anywhere at a Personal Level. In addition to that its subcomponents may have much great applications as stated.

- A new Improved Gesture recognition method is implemented in this application and it can be used in any other gesture recognition Software.
- With the Help of this application we can make public announcements, conferences, news for Deaf and dumb Community.
- As blind have a communication medium in form of software, our concept is an addition for Deaf as well as Dumb people as communication medium.

VIII. CONCLUSION

In this paper, we stated the development of an interpreter for communication between Normal Person and a Deaf and Dumb Person, for the individuals of a country India where it holds a great Importance because of higher population and less development in respected field.

Talking in technical terms for development of such an idea to a realistic model requires heavy processing of Sentences and Gestures is required over a network to have an efficient result. The subcomponents, Natural Language Processing responsible for Translation of ISL to English and via versa. Gesture Recognition is responsible for recognition by a unique and more efficient way required for Gestures Input. The whole set of modules contribution created an excellent application which will defiantly bridge the gap between the Normal Person and Deaf and Dumb Person and give the chance to the disable Person to not only stand but race in today's world. It is a step ahead in the direction of a new Era where Disability has no Place.

REFERENCES

- [1] Rostislav Miarka and Martin Žáček "Knowledge patterns in RDF graph language for English sentences" in Proceedings of the Federated Conference on Computer Science and Information Systems pp. 109–115 ISBN 978-83-60810-48-4
- [2] UMESH CHANDRA JAISWAL, DR. RAVINDRA KUMAR and DR SUSHIL CHANDRA "A Structure Based Computer Grammar to Understand Compound-Complex, Multiple-Compound and Multiple-Complex English Sentences in 2009 International Conference on Advances in Computing, Control, and Telecommunication Technologies
- [3] Xiaohui Shen, Gang Hua, Lance Williams and Ying Wu "Motion Divergence Fields for Dynamic Hand Gesture Recognition"
- [4] T.F. Michael Raj, Durga Shankar Baggam, M. Prabhu and H. Manikandan "A Domain Based Approach for the Animated Cum Automated Model for Deaf and Dumb Using NLP" in Research Journal of Applied Sciences, e-ISSN: 2040-7467
- [5] Rada Mihalcea¹, Hugo Liu², and Henry Lieberman² "NLP (Natural Language Processing) for NLP (Natural Language Programming)"

- [6] J. Ignacio Serrano and Lourdes Araujo "Evolutionary Algorithm for Noun Phrase Detection in Natural Language Processing" in TIC2003-09481-C04 and FIT150500-2003-373.
- [7] Poonam Suryanarayan and Anbumani Subramanian, Dinesh Mandalapu "Dynamic Hand Pose Recognition using Depth Data" in 2010 International Conference on Pattern Recognition
- [8] A. M. Turing "COMPUTING MACHINERY AND INTELLIGENCE" in A. M. Turing (1950) Computing Machinery and Intelligence. Mind 49: 433-460.
- [9] Md. Abu Nuser Musud', Md. Muntusir Mamun Joarder', Md. Turiq-UI-Azam "A general approach to natural language coersion"