Machine Learning Oriented Gesture Controlled Device for the Speech and Motion Impaired

Ashish Gupta
Dept of Electronics and Communication Engineering
JSS Science & Technology University
Mysuru, Karnataka, India – 570006
ashishgupta63966@gmail.com

Sandesh Jagadish
Dept of Electronics and Communication Engineering
JSS Science & Technology University
Mysuru, Karnataka, India – 570006
sandeshjagadish@gmail.com

***Abstract*— India is an extremely populous country with a massive population of 1.2 billion residents, and a large proportion of people with disabilities in both rural and urban India. Speech impairment is common with people with hearing loss since birth. Among the total disabled population, about 27% have movement constraints and hence are confined to wheelchairs. This paper proposes the idea of using machine learning implementation to develop a device that can benefit the speech and motion constrained population. Gesture control plays an essential role in order to convert the sensor data into speech output or operating a pick and place bot for the motion constrained. Various algorithms are interfaced with the device to provide efficient functionality and throughput. Machine learning and data analytics technologies have been on the rise recently and are finding applications in various domains and industries.**

Keywords—Gesture, neural networks, speech, motion, wireless.

I. INTRODUCTION

In a country like India, where unemployment and poverty are major concerns, the disabled population is adversely affected. Bridging the gap with the disabled and to bring them on par with us, is an essential step which India requires, and is not an easy one. The need for an innovative solutions which are accessible without complications, at cheap prices is more important than ever if India has to progress to become a superpower in the coming years. Our product Gesture Konnect is designed for the applications which could benefit two types of disabilities- speech and movement. The following graph shows the percentage of people who are affected by such disabilities.

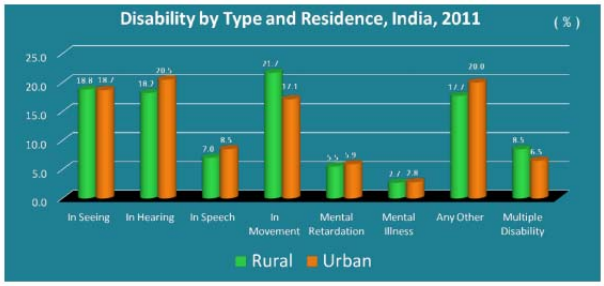

Figure 1: Statistics relating to the disabled population in India as of 2011

There are about 21.7% of total rural disabled population and around 17.1% of total urban disabled population having motion constraint. This population is neglected and underestimated in this fast paced era. Through this project, we intend to provide them with the ability to be a part of the growth in this developing nation.

For this application, the gesture controlled pick and place assistant uses a camera feed for the user to navigate and the gesture performs the picking and placing operations of objects. This creates an extremely useful application for the movement-impaired, wheelchair users and for the elderly people. The lack of strength in movement will force them to use an easier approach to pick and place objects using a simple methodology like gestures.

II. PROPOSED DESIGN OF THE DEVICE

There are many sign language translators in use at present but the proposed device addresses an effective approach to provide the disabled with means to communicate like normal people and perform everyday activities. The device referred as 'Gesture Konnect' uses a modular compact design to implement this concept. The focus is on two major disability issues - speech and motion. The functionality comprises of multiple modules interfaced for coherent operation.

A. GestureKonnect for the speech disabled

The proposed device is designed to be a wearable device like a glove that is worn on the hand of the person. A compact microcontroller forms the brain of the transmitter system. The sensing unit is placed at specific nodes of the glove to record the position of the gesture. Accelerometer or gyro sensors can be deployed easily. The sensing data is collected by the microcontroller which transmits it to a new microcontroller via the RF/Bluetooth transceiver attached to the system.

The received data is stored in the second microcontroller which forms the test dataset. The training dataset undergoes the neural network LVQ algorithm technique for multinomial clustering. This is implemented initially in a separate system. The cluster representing the gesture is matched with the test data. Each cluster is a character. Multiple gestures result in a set of characters or string.

A spell check algorithm is used to enhance the string to form a coherent word by determining the most probable word related to the multi-gesture characters collected. The words are fed into a text-to-speech conversion unit where a computer generated voice is used to output the words.

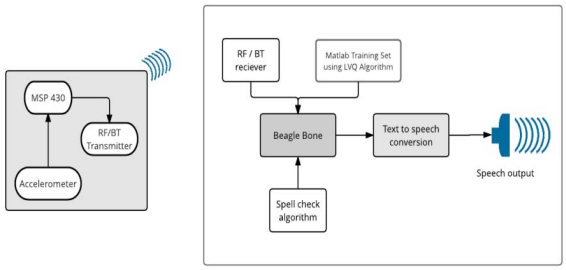

Figure 2: Schematic representation of GestureKonnect for the speech disabled

B. *GestureKonnect for the motion disabled*

The device is fitted with two modes of operation for dual application. The movement impaired and elderly people can make use of this operation to pick and place objects with ease. The gestures are generated from the sensing unit like accelerometers and the data is sent to a microcontroller, which transmits it to the bot having the second microcontroller. The second microcontroller receives this data through RF/ Bluetooth transmission.

The gestures are mapped to different tasks and motor drivers to control the movement and the pick and place operation of the bot using Learning Vector Quantization (LVQ) artificial neural network.

.

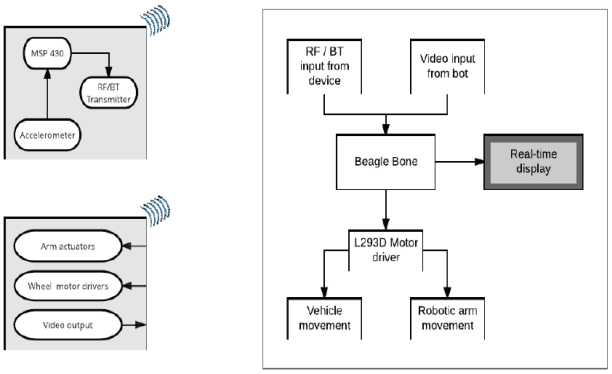

Figure 3: Schematic representation of Gesture Konnect for the motion impaired.

Motor drivers are used to control the motion of the bot and the servo motors in the arm for picking and placing of objects. A live video feed from a camera module placed with the bot is sent to a real time display unit which is held by the user for better navigation and control of the bot from a large distance.

III. WORKING PRINCIPLES

The functioning of the product is controlled by 3 fundamental concepts. These techniques have been optimized to maintain the overall desired efficiency and fidelity.

A. *Gesture recognition with neural networks*

An artificial neural network (ANN) Learning Vector Quantization (LVQ) algorithm is used to train and recognize gestures from GestureKonnect. The algorithm is a prototype based supervised classification algorithm which is represented by prototype weights which are defined in the feature space of the observed data.

The data here is obtained from the x-y-z coordinates of the accelerometer sensor in the wearable device. These coordinates form the input vector. The sensor readings are initially trained using MATLAB by defining the known categories, in this case, the characters and bot movements. The neural network consists of m number of neural output elements each having a weight W=(w(i),…..,w(n)). The input list X is used to train the network with a learning rate ‘a’.

Let the input vector x=(x,y,z) from the accelerometer and the weight of the jth neuron wj=(w1j,w2j,w3j) be considered, the category Cj is represented by the jth neuron, which is pre-assigned. Let T be the correct category of the input during training. The Euclidean distance between the input vector and the weight vector of the nth neuron is given as:

$$D(j) = \sqrt{\sum_{i=1}^{n} (x_i - w_{ij})^2}$$

The weight vectors and the learning rate ‘a’ is initialized. For each input vector, the j^{th} neuron is determined so that D(j) is minimum.

If $T=C_j$, then

$$w_{j(new)} = w_{j(old)} + a \,.\, (x - w_{j(old)})$$

that is, move the weight vector w toward the input vector x

If $T!=C_j$, then

$$wj(new) = w_{j(old)} - a \,.\, (x - w_{j(old)})$$

that is, move the weight vector w away from the input vector x

This is done for a fixed number of iterations till the weights are optimized and represented uniquely in a cluster with respect to a particular character or task.

This is described in the following images. Note how the weight for each cluster of data moves with respect to the input data set of the same category as the training progresses.

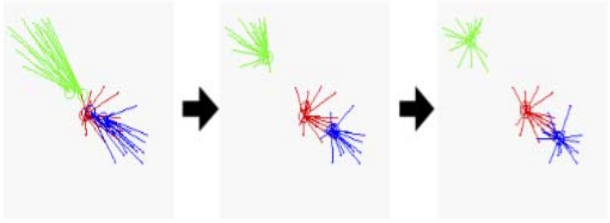

Figure 4: Learning Vector Quantisation (LVQ) algorithm representation

B. *Fuzzy String Matching*

To enhance the ease and efficiency of the translation, we have chosen to use Fuzzy String match algorithm. This technique matches a pattern approximately, rather than exactly. This in turn minimizes the errors during conversion, and also reduces the number of characters required to represent a word, with appreciable accuracy. The difficulty of approximate string matching is divided into two sub-problems: finding approximate match for the substring inside the given string, and finding the dictionaries that match approximately to the string.

The closeness is measured in the number of fundamental steps required to convert the pattern into an exact match. This is termed as the edit distance between the pattern and the string.

Some primitive steps are
Deletion - b e a t → b * a t
Insertion - b * a t → b e a t
Substitution - b e a t → b o a t

The process of matching the pattern involves substitution of characters, or swapping of characters within the pattern. To keep a track of the changes, the algorithm uses a single unweighted cost factor, which refers to the number of primitive steps required to make the match. Matches with lower cost factor are given more importance.

With the obtained pattern $P = p1\ p2\ p3....pm$ and the text strings data set $T = t1,t2,...,tn$, the algorithm outputs the set $R = ti,...,tj$ from the set T, which have the least edit distance from the patterns P.

For every position j in the text T, and every position i in the pattern P, traverse through all T substrings ending j, and list the substrings with least edit distance to the first i characters in pattern P. Taking this this minimal distance as E(i, j) and computing it for all i, j, the substring for which E(m, j) is minimal is the required match for the input pattern(m being the pattern P length).

C. *Text to Speech (TTS) Conversion*

Text to speech synthesizer converts any input text into an audio format and reads it aloud. The application of TTS here is to convert the strings after the matching, and play them through the speaker to convey the user's message to the listener. The operation of TTS is performed by 2 sub tasks - the Natural Language Processing (NLP) and the Digital Signal Processing (DSP). The NLP synthesizes a phonetic transcription of the text read, based on the desired rhythm and intonation.

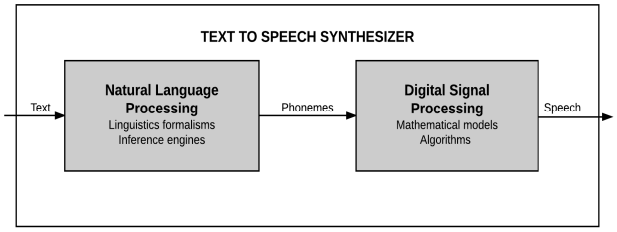

Figure 5: Block diagram of TTS synthesizer

The NLP block is mainly responsible for letter-to-sound and prosody generation. It also consists of a morpho-syntactic analyzer, emphasizing the need for syntactic processing in a high quality TTS system.

The operations of the DSP block involve the computer analogue of controlling the articulatory membranes and the frequency of vibration of the vocal slits. This to ensure the matching of required intonation. These actions are guided by articulatory constraints, since the phonetic transitions are more essential than a stable state for accurate understanding of speech.

ACKNOWLEDGMENT

We are grateful to Prof. M N Shanmukha Swamy (JSS Science and Technology University), Prajwal Kashyap (JSS Science and Technology University), Ritwick Medikeri (University of Michigan, Ann Arbor), and Prof. Sriram Kalyanaraman (University of Florida) for helpful reviews on the draft of this paper and for their valuable suggestions. We also extend our thanks to Saurabh NK ((JSS Science and Technology University) for proof-reading and editing the grammatical errors in the paper.

FEASIBILTY

The proposed product is designed with the motto to achieve high accuracy with reasonable cost for implementation. The analysis is done using a pragmatic approach and to make the product accessible to the target audience effortlessly. The idea is developed into a product keeping in mind all the issues that could arise and once the product is developed, it is subjected to testing and customer review and feedback.

The efficiency of gestures to perform actions is dependent on the machine learning algorithm. Larger training datasets are used to improvise the parameters of the learning model to achieve higher mapping of tasks to gestures. The prototype provides the necessary testing and quality assurance of the product which is subjected to the new test dataset directly obtained from the user.

CONCLUSION

Conventionally, the mute population uses sign language to communicate which may be hard to perceive and interpret for some and definitely impossible to those who do not know the sign language. The product designed in our solution uses a sophisticated machine learning neural network to determine the speech generated from gestures provided by the deaf-mutes. This control provides efficient task performance with limited components and results in an increase of durability and reliability. This product is small and user friendly and the operation can be explained to the user with ease. Since the gestures are letter/character oriented and the spellcheck algorithm tries to rectify and complete the character set to form meaningful words, this allows the deaf-mute person to form a variety of words and thus improving their ability to communicate. Portability is another advantage which could convert a deaf-mute into a normal moving talking person.

This product can be extended to other applications like to control a bot to pick and place objects for the wheelchair confined, movement-impaired and the elderly, a camera module could provide a live feed to the user to operate this machine. The machine learning algorithms to analyse gestures increase the level of accuracy and provide for the user to perform tasks efficiently.

REFERENCES

[1] Michael Biehl, Anarta Ghosh , "Dynamics and Generalization Ability of LVQ Algorithms" , Institute for Mathematics and Computing Science University of Groningen P.O. Box 800, NL-9700 AV Groningen, The Netherlands, Barbara Hammer Institute of Computer Science Clausthal University of Technology D-38678 Clausthal-Zellerfeld, Germany (2007)

[2] Atsushi Sato & Keiji Yamada, "Generalized Learning Vector Quantization" , Information Technology Research Laboratories, NEC Corporation 1-1, Miyazaki 4-chome, Miyamae-ku, Kawasaki, Kanagawa 216 .

[3] Endarapu Vanitha, Pradeep Kumar Kasarla, E Kumaraswamy on ” Implementation of Text-To-Speech for Real Time Embedded System Using Raspberry Processor ”.

[4] P.V.N. Reddy , “ Text to Speech Conversion Using RaspberryPi for Embedded System “, Professor, Department of ECE, S V College of Engineering, Tirupati, A.P, India.

[5] Baeza-Yates R, Navarro G , "A faster algorithm for approximate string matching".