International Journal of Advanced Technology in Engineering and Science
Vol. No.08, Issue No. 03, March 2020
www.ijates.com
ijates
ISSN 2348 - 7550

SMART ASSISTANCE FOR - MUTE CARE TAKER

Dr.C.VENKATESH[1] ,G.SWATHI[2], S.SARULATHA[3],
S.PRIYADHARSHINI[4], S.SOWNDHARYA[5]

[1]PROFESSOR/ECE, SENGUNTHAR ENGINEERING COLLEGE, INDIA

[2, 3, 4, 5]FINAL YEAR STUDENTS /ECE, SENGUNTHAR ENGINEERING COLLEGE, INDIA

ABSTRACT:

In order to develop a Voice Assistance System for the dumb people, several voice samples are to be collected to train the software model. The voice assistance system is for the dumb people who can't speak like a normal people. The system is created by combining the four algorithms called Machine learning(Deep speech recognition)algorithm, Connectionist Temporal Classification(CTC) algorithm Language model(LM)algorithm, Text To Speech (TTS) algorithm. The system works under the principle of deep speech recognition. First the model is trained with dumb people voice signal. Then the model stores the weights and bias for all the training data. Then this model is ready to get and process the input (dumb people voice signal). It produces output as text format and the proper prestored voice signal for the given particular input(dumb people voice signal).

Key words: *ML,CTC,LM,TTS, Training, Text format, Prestored voice signal.*

- The Introduction part is described in section I.
- The literature survey is described in section II.
- The Proposed System method is described in section III.
- The Step by Step installation method is described in section IV.
- The Machine Learning Algorithm is described in section V.
- The Connectionist Temporal Classification (CTC) is described in section VI.
- The Language modeling(LM) is described in section VII.
- The Text to Speech Converter(TTS) is desceibed in section VIII.

I. INTRODUCTION:

The big disability of the dumb people is that they can't speak like a normal person. They depend on hand gestures to communicate with others. The existing system is only useful for the people who mutilated after birth because of some accidents. By the proposed system, the dumb people by birth, can overcome their disability. For example when a dumb people trying to say hello by his voice, The Voice assistance system will take the voice signal and produce text format for hello and a voice signal of hello. By this text format dumb people can easily communicate with deaf people. By the voice format dumb people can easily communicate with normal people.

II. LITERATURE SURVEY:

- **An innovative communication system for deaf ,dumb and blind people**
  [https://www.scribd.com/document/424696390/An-innovative-communication-systemfor-deaf-dumb-and-blind-people]
  It provide technique for a blind person to decipher a text and it can be achieved by capturing an image through a camera which converts Text to Speech (TTS). It provides a way for the deaf people to read a text by Speech to Text (STT) conversion technology.Also it provides a technique for dumb people using text to voice conversion system-using-raspberry-pi]

- **A novel technique for speech recognition and visualization**
  [https://www.researchgate.net/publication/325359086_A Novel Technique _for_Specch Recognition_and_Visulatization_Based_Moblie_Applicationn_to_Support_Two-Way_Communication_between_Deaf-Mute_and_Normal_Peoples]

- In this system, Mel Frequency Cepstral Coefficients (MFFC) based features are extracted for each training and testing sample of Deaf –mute speech. The hidden markov model toolkit (HTK) is used for the process of speech recognition. The application is also integrated with a 3D avatar for providing visualization support

- **Hand gesture recognition and voice conversion for speech impaired [https://www.ijert.org/research/hand-gesture-recognition-and-voice-conversion-system-for-specch-impaired-IJRTCONV5IS01047.pdf]**

  This proposed a new technique called artificial speaking mouth for dumb people. This system is based on motion sensor. For every action the motion sensor get accelerated and give the signal to the microcontroller. The microcontroller matches the gesture with the database and produces the speech signal. The system also includes a Text To Speech conversion (TTS) block that interprets the matched gestures.

III. PROPOSED METHOD:

System that act like an interface between dumb people and normal people. Software model should be created by combining all the four algorithms namely ML,CTC,LM,TTS.Software model should be trained with the dumb people voice datasets. Theweights and bias are stored in the software model. In the ML, deep speech recognition algorithm is used. CTC which uses 3000 hours normal people voice signal for predicting the correct word.Language modeling which checks the spelling of the word.TTS which converts the text into speech. The software model gives the output as text and voice signal for the input voice signal of dumb people

BLOCK DIAGRAM

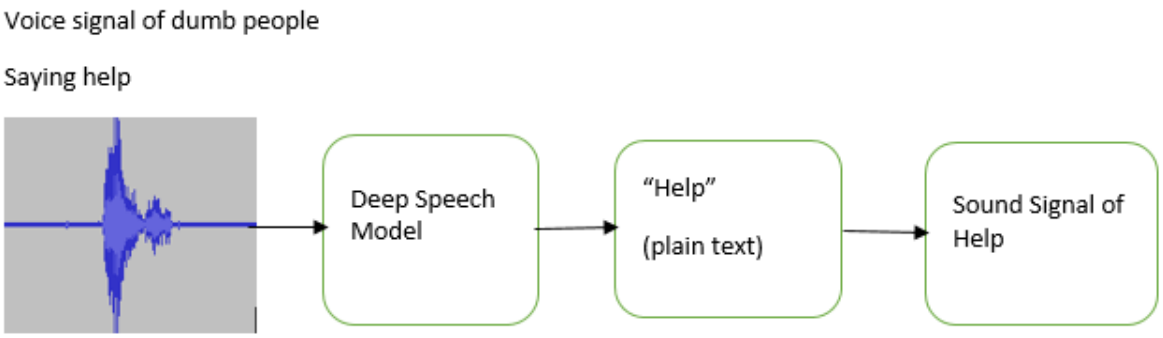

Fig 1-Speech Recogintion model

In Fig -1 describe how dumb people speech signal convert to plain text and sound signal using CTC algorithm ,LM algorithm and TTS algorithm.

AI in Speech Recognition

Speech recognition is the ability of a machine or program to identify words and phrases in spoken language and convert them to a machine-readable format.

IV.PROJECT STEP BY STEP INSTALLATION:

1. Create and activate a virtual environment

virtualenv -p python3 env

source env/bin/activate

2.Deep speech github:

git clone https://github.com/mozilla/DeepSpeech

cd DeepSpeech/

git checkout v0.6.1

pip3 install $(python3 util/taskcluster.py –decoder)

pip3 install -requirements.txt

3.English pretrained model checkpoints:

Gotohttps://github.com/mozilla/DeepSpeech/releases

V. Machine learning algorithm:

In the proposed system supervised algorithm is used. Supervised algorithm refers to the technique of labelling the all related contents of single data. Here the text format of the word spoken by the dumb people is being taken as a label context.

The voice by signal of dumb people is taken as input and the features namely, speech length, voice pattern, tone frequency are being extracted the windowing technique.

Tensor flow frame work: It is an open source artificial intelligence library, using data flow graphs to build models. It allows developers to create large-scale neural networks with many layers. In the proposed system six layers are used. Tensor flow allows to build a neural network models to recognize spoken words. These models typically use the Recurrent Neural Network (RNN).

Programming language: Python.

Speech Recognition Architecture: Deep speech architecture.

Architecture:

In below fig-2 describes deep speech architecture is having 2048 neuron and 5 hidden layers. Thus it provide a good training space(environment) to the software model. Audio is sampled at 44.1 kHz (44,100 readings per second). But for speech recognition, a sampling rate of 16 kHz (16,000 samples per second) is enough to cover the frequency range of human speech**.**

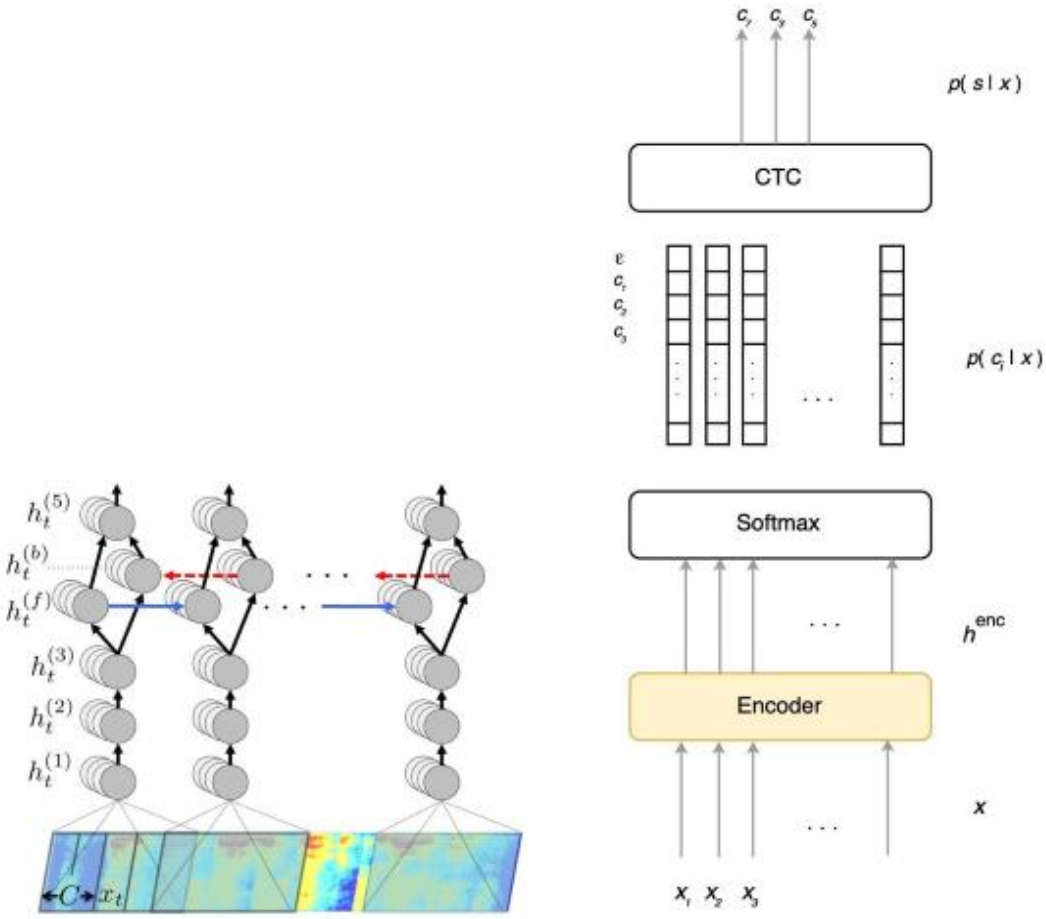

Fig-2 Deep Speech Architecture

Speech recognition software uses natural language processing (NLP) and deep learning neural network to break the speech down into components that it can interpret. It converts these components into a digital state and analyzes segments of content. The software trains on a data set of known spoken words or phrases and make predictions on the new sounds, forming a hypothesis about what the user is saying. It then transcribes the spoken word into text.

VI. Connectionist Temporal Classification (CTC):

[NCILC-2018_paper_4pdf]

CTC is a type of neural network output and associated scoring function, for training Recurrent Neural Network (RNN) such as LSTM networks to tackle sequence problems where the timing is variable. It uses 3000 hours continuous English speech content to compare the letters generated from ML algorithm

and to produce word format. Thus it produce the word format latter this word undergoes spell check in the LM algorithm.

[https://towardsdatascience.com/hello_world_in_specch_recognition_b2f43b6c5871]

VII. Language modeling(LM):

Skip gram:

A language model learns to predict the probability of a sequence of words. In the proposed system skip gram algorithm is used and kenlm language model toolkit is used to build the model.

Word representation represents the word in vector space so that if the vector are close to one another means that those word are related to one another.As the vocabulary of any language is large and cannot be labeled by human and hence we require unsupervised learning techniques that can learn context of any word on its own. Skip gram is one of the unsupervised learning techniques used to find the most related word for a given word.

Variable used:

The dictionary of unique words present in the data set or text. This dictionary is known as vocabulary and is known words to the system. Vocabulary is represented by 'V'.

N is the number of neurons present in the hidden layer.

The windows size is the maximum context location at which the words need to the predicted.

The windows size is denoted by C.

The dimension of input vector is equal to |V|.

The weight matrix for the hidden layer (W) is of dimension [|V|,N] which return size of an array.

The output vector of the hidden layer is H [N].

The weight matrix between the output layer(W') is of dimension [N,|V|].

The dot product between W' and H gives an output vector U[|V|].

[https://towardsdatascience.com/skip-gram-nlp-context-words-prdiction-algorithm-5bbf34f84e0c]

VIII. TTS:

Text to Speech (TTS) converter:

[https://heartbeat.fritz.ai/a-2019-guide-to-specch-synthesis-with-deep-learning-630afcafb9dd]

This part describes neural network architecture for speech synthesis directly from text. The system is composed of a recurrent sequence-to-sequence future prediction network that maps. Character embeddings to mel-spectrograms. Using modified network mel-spectrograms converted to speech.

Spectrogram prediction networks:

The network is composed of an encoder and decoder with attention. The encoder converts a character sequence into a hidden features representation with the decoder consumes to predict a spectrogram.

Modified wave net:

Wave net is a deep neural network for a generating raw audio. It's able to generate relativity realistic sounding human (Voice) by directly modeling waveforms using a neural network method trained with recording or real speech.

IX. REFERENCES:

1.https://heartbeat.fritz.ai/the-3-deep-learning-frameworks-for-end-to-end-speech-recognition-that-power-your-devices-37b891ddc380[https://www.ijraset.com/]

2.International Journal for Research in Applied Science and Engineering Technology (IJRASET), ISSN: 2321-9653

3.WHO, "Deafness and hearing loss," Fact sheet, vol.3006', 2017

4. International Journal of Engineering Research and Technology (IJERT),ISSN:2278-0181,ICIATE-2017 Conference proceedings

5. https://distill.pub/2017/ctc

6.https://medium.com/@ageitgey/machine-learning-is-fun-part-6-how-to-do-speech-recognition-with-deep-learning-28293c162f7a

7.Kalaimagal, M., Meerasri, M., Murugesh, M., Thirunavukkarasu, M. and Venkatesh, C. (2015). Investigation of Human Behavior using Biometrics. International Journal of Science Technology & Engineering, 1(9), pp.19-23.

8.P.Ponmurugan, B.Priyadarshini, P.Preetha, V.Preethikadevi, R.Divya, "Health Care Assisting Chatbot for symptoms and dosage prediction using IoT", REST Journal on Emerging trends in Modelling and Manufacturing, Vol. 4, No.2, 2018, pp.50-54.

9. https://github.com/**mozilla**/**DeepSpeech**/**releases**/**download**/v0.6.1/**deepspeech**-0.6.1-**checkpoint**.tar.gz

10.https://en.m.wikipedia.org/wiki/connectionlist_temporal_classification

11.https://towardsdatascience.com/skip-gram-nlp-context-words-prediction-algorithm-5bbf34f84e0c