

# Quantum Speech

*Speech Analytics with Emotions Recognition  
for AWS F1/FPGA/ASIC/Cloud Based Call (Contact) Centers*

## Common Objectives:

To develop and to promote use of Multi-Lingual Emotions Recognition AI Models-AI Adapters under common banner of Quantum Speech consisting of Emotions Recognition and Speech Analytics through public participation for public use. Quantum Speech is a common platform (as well as collection of AI Models-AI Adapters for different natural languages) where we showcase Emotions Recognitions AI Models-AI Adapters for participants and the participants showcase different Speech Analytics Software (built using the AI Models-AI Adapters) for others.

## Motivation:

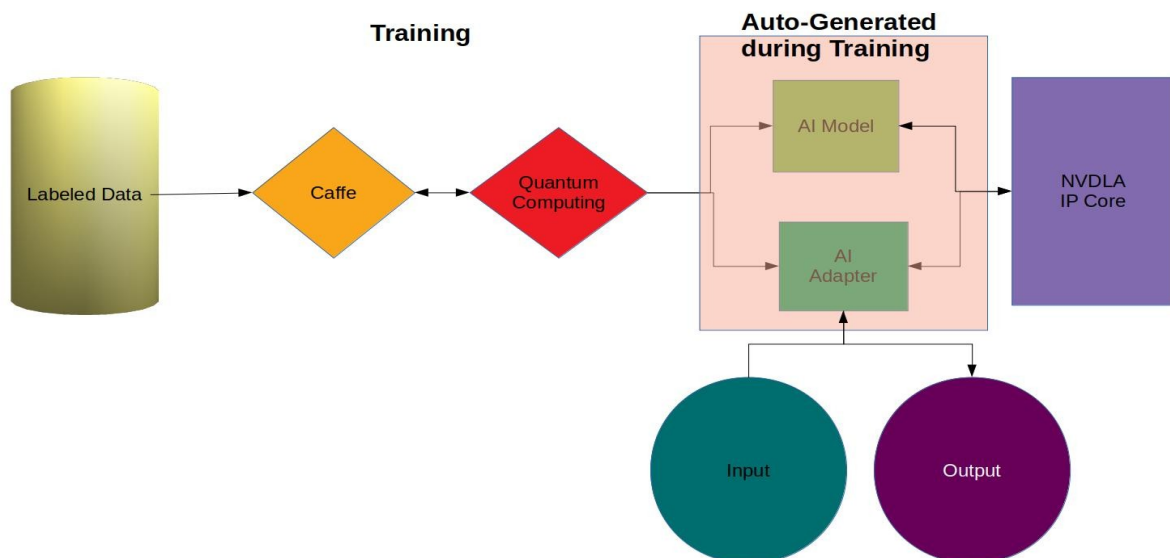
Nature of emotions like as superposition states as shown in the Plutchik's human emotion model given below and nature of chain of emotions in speech like as entangled states motivate to treat emotions Recognition AI models development for different natural languages with the help of brain like quantum computing environment.

## Plan:

Initially the public participation will be in audio data labeling and speech analytics software development usually developed using database and business intelligence software, but other more advanced areas like Chisel/Scala/Verilog/RISC-V may be opened in future as per participants demand if the initial response is good. The work coordination will be done through github.

**Quantum Speech** consists of series of proposed deep learning multi-lingual emotions Recognition models super-refined by Quantum Computing and auto-generated by our Quantum AI Training and AI Model-Adapter Generation Process given in the diagram below and speech analytics software developed by the participants/developers:

### Quantum AI Training and Model-Adapter Generation Process



**The above process can also generate AI Models-Adapters for different fields of AI applications other than emotions Recognition.**

**A. Labeled Data:** Only the label data preparation process for emotion Recognition is described here though the process which can accomodate labeled data from other artificial intelligence application areas as well. The participant observes emotions word by word labels these according to the naming convention gievn below and our training process generates learned AI-Models and AI-Adapters for emotions Recognition.

We can use Audacity for labeling data and VLC to record audio files from downloaded youtube videos for further processing though Audacity, please take care of copyrights. Links are given below to help in label data generation:

Download Audacity:

<https://sourceforge.net/projects/audacity/>

Tutorial on Labeling Data with Audacity:

<https://www.youtube.com/watch?v=JR7ehNSqIq4>

Download VLC Media Player:

<https://www.videolan.org/vlc/index.html>

Tutorial on Downloading Youtube Videos with VLC:

<https://www.youtube.com/watch?v=9G2VpHhPTgM>

Converting Video to Audio Files with VLC:

[https://www.youtube.com/watch?v=Zmi3-a\\_sD7s](https://www.youtube.com/watch?v=Zmi3-a_sD7s)

Many tutorials on denoising and preparing audio files before labeling with Audacity are available on the internet. Please ensure to include mutliple and different sampled labeled data files for as many words as usually used to descibe different emotions in a given natural language are covered. Please use Plutchik's simplified human emotion model given in this document to describe emotions with the ultimate aim that the Plutihik's detailed human emotion model will be used in generated AI Model for emotions Recognition once sufficient labeled data is made available over the time in the future versions.

### **Labeled Data File Naming Convention Compatible with the Training Process:**

Each labeled file should have a unique filename. The filename should consist of seven two-digit numerical identifiers, separated by hyphens (e.g., 02-01-06-01-02-01-12.wav). Each two-digit numerical identifier should define the level of a different experimental factor. The identifiers are ordered: FileType–Channel–Emotion–Intensity–Statement–Repetition–Actor.wav. The numerical coding of levels is described in following table For example, the filename “02-01-06-01-02-01-12.wav” refers to: Video-only (02)–Speech (01)–Fearful (06)–Intensity normal (01)–Statement “Hello World” 02)–First repetition (01)–12<sup>th</sup> Actor, female (12). The gender of the actor is coded by the actor's number, where odd numbered actors should be male, even numbered actors should be female.

Identifier	Description
<b>FileType</b>	01=Audio-Only, 02=Video-Only, 03=Audio-Video
<b>Channel</b>	01=Speech, 02=Song
<b>Emotion</b>	00=Normal, 01=Angry, 02=Afraid, 03=Surprised, 04=Happy, Sad=05, Disgusted=06
<b>Intensity</b>	01=Normal, 02=Strong
<b>Statement</b>	01="Hello, World" 02-"How Are You", 03=...
<b>Repetition</b>	01=1 <sup>st</sup> , 02=2 <sup>nd</sup> , 03=3 <sup>rd</sup> , ...
<b>Actor</b>	01=Urvi, 02=Neha, 03=Priya, ...

## B. Caffe:

Caffe is a deep learning framework made with expression, speed, and modularity in mind. It is developed by Berkeley AI Research (**BAIR**) and by community contributors. **Yangqing Jia** created the project during his PhD at UC Berkeley. Caffe is released under the **BSD 2-Clause license**.

For more details, please visit: <https://caffe.berkeleyvision.org/>

## C. Quantum Computing (QC):

Quantum Mechanics (QM), a branch of physics related to microscopic world of electrons, atoms, molecules etc has ushered the revolution of semiconductors and consequently revolutionized IT Industry and Telecommunication Industry to name a few apart from Chemical and Biochemical Industries. The mathematics behind the Quantum Mechanics (QM), which can also be summed up as Quantum Computing (QC) in the information science, is the same as believed to be behind the neural processing of the biological/biochemical brain, though the links, in physics, behind the two are yet to be clearly understood by the scientists. This is just in confirmation with physics-mathematics traditions-some times physics leads mathematics and sometimes mathematics leads physics.

QC is quite different from the classical computing with respect to flow bits (qubits in QC)-details of QC will quite justify a thousand page plus volume itself so we will focus on implementations of it with respect to speech only. Speech not only convey signals to be translated into text and understood, but the complete understanding of speech involves emotions apart from text. Since speech containing emotions can be better processed using QC in brain like processing environment so the motivation for emotions Recognition aspects of Quantum Speech is self evident.

QC is yet to be available as a commodity product for masses. Many commercial setups for QC algorithm development and experiments are available at affordable cloud platform open for public use, for example:

IBM Q Experience:

<https://quantum-computing.ibm.com/login>

Microsoft Quantum Computing:

<https://www.microsoft.com/en-in/quantum/>

D-Wave Quantum Computing:

<https://www.dwavesys.com/quantum-computing>

With the availability of above the dawn of QC has arrived and is real, which will pave the way for mass applications in near future.

Though our QC emulation system can be easily ported to above such systems but we comfortably use our QC emulation system based on Apple Mac Clusters and Xilinx Zynq FPGA Boards without significant loss of comparable speed and comparable accuracy at comparably much cheaper price with comparably significantly higher availability time and controlled environment. The process described in the above diagram is made available on our QC emulation system and above all is available free of cost to those who participate with us in development of quantum speech under free of cost open source licensing scheme of their choice where the AI Model will be available under open source license and the relevant AI adapter will be distributed free of cost in binary form as shareware. Commercial options can be worked out in consultation with the concerned participant of Quantum Speech for a particular language(s) if the participant so desire.

#### **D. AI Model and AI Adapter:**

Since QC is used here for super-refining the AI Models and AI Adapters Generation during training process, so no attempt is made to implement of QC at FPGA/ASIC level or at AWS F1, the generated AI Adapters automatically take care of the necessary minimised and optimised AI implementation side requisits in classical computing itself. AI Models are basically NVDLA compatible data models.

#### **E. NVDLA:**

NVDLA , the FPGA/ASIC IP Core ,is provided by NVIDIA under free of costs open source for commercial use. The IP Core can be freely used in FPGA/ASIC/AWS F1 implementations and can be very well integrated to RISC-V architecture.

For more details, please visit:

<http://nvdla.org/>

<https://riscv.org/>

<https://aws.amazon.com/ec2/instance-types/f1/>

#### **F. Input and Output:**

Live audio stream usually accompanied by pre-processed text by pocketphinx speech Recognition software is transferred to the AI Adapter which in turn hands it over to NVDLA and the AI Adapter collects the results with recognized emotion in text forms suitably appended to the speech text recognized by the pocketphinx for further processing by chatbot like E.D.D.I and then there after festvox text to speech converter. At another end AI Adapter can dump speech data directly into database for further processing and visualization through a business intelligence software.

For more details, please visit:

<https://cmusphinx.github.io/wiki/>

<http://www.festvox.org>

<https://eddi.labs.ai/>

## **G. Copyrights:**

As stated in the above diagram, Charu Bhardwaj and Indu Mitra own the Quantum AI Training and Model-Adapter Generation Process. Since the labeled data will be owned and developed for a particular spoken language by the participants so the relevant contributed AI Model (under the Quantum Speech) copyrights will be shared with the contributing participants of that particular AI Model under the terms and conditions framed in advance in consultation with the participant. The relevant AI Adapter will be available as shareware in executable binary form if the participant agrees to release the AI Model free of cost to public.

## **H. Contact Details:**

Concerned Person: Charu Kanzehr Bhardwaj  
Email: [charu.k.bhardwaj@gmail.com](mailto:charu.k.bhardwaj@gmail.com)

Any query or suggestion related to further improvements may be sent to the above email address.

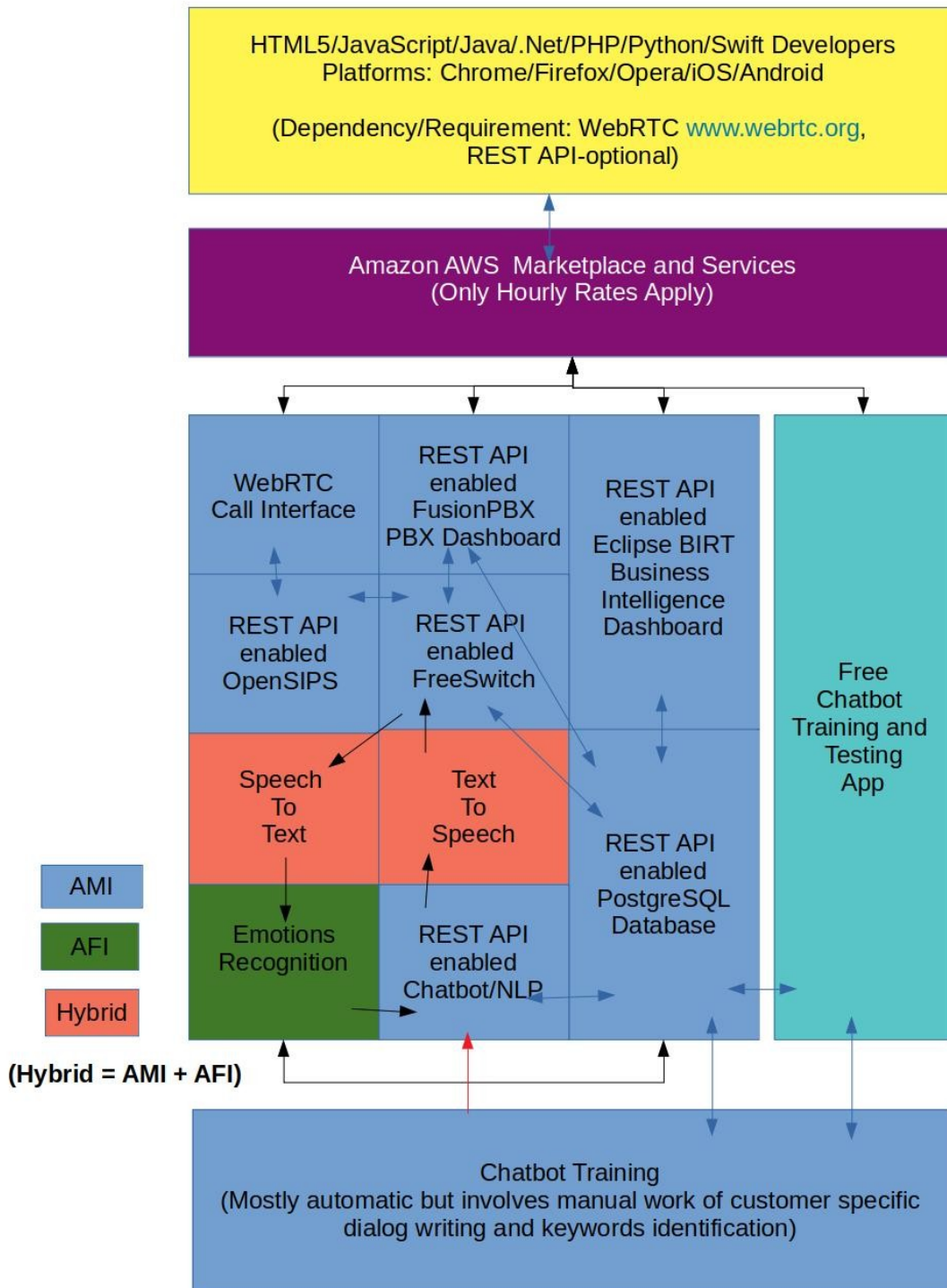
## **Execution/Implementation:**

In order to promote Quantum Speech, we see, as first instance, may be due to our past experience, great market in Call/Contact Industry for Emotions Recognition enabled Speech Analytics, though there are other significant applications in other market areas as well like toys, Alexa like speakers etc. For your reference, we propose following techno-commercial model for Call/Contact Center Industry with user required scalability, adjustments and configurations. It can be as small as WebRTC enabled voiced chatbot or as scalable as Twilio like virtual telecommunication operator.

The techno-commercial model given here is meant to be executable on Amazon AWS F1 instances. We have developed Quantum Speech Simulator containing Emotions Recognition, Text-to-Speech, Speech-to-Text (excluding other components mentioned in the techno-commercial model) along with AI Model and AI Adapter in Amazon Linux 2 VDI file which can be run under Virtual Box. In order to make on-premise testing and validation without paying for Amazon AWS F1 services possible for participants, the simulator along with your AI Model and our AI Adapter (in VDI file) can be made available to the participants free of costs. Practically speaking, participants need to build the labeled data at their own costs and rest is available free of costs before they decide to go to commercial/production implementation at Amazon AWS F1.

AI Adapter running under simulator can be connected to the database, this feature can be particularly helpful for those participants, who want to do further software development like Speech Analytics or test deployment at their own premises.

## Techno-commercial Model





## A. Applications:

Virtual-Agents & Speech Analytics for Marketing/Contact Centers, Personal Assistants, Robotics, Toys/Virtual Pets, AR/VR/MR/Games, Automated Car Voice Communication, Autopilot Voice Communication in Civil Aircrafts, Fire-control Voice Communication in Military Aircrafts, Medical/Forensic/Police-Inestigation which can be further enhanced by Vision, EEG, ECG, Ultrasound, MRI inputs etc.

## B. For Customers:

Analysis of emotions in speech is very important for marketing/contact centers and other applications mentioned above, advent of virtual agents in contact centers has further increased its importance. Plutchik's human emotion model as shown in detailed and simplified diagrams below:

### Detailed Model:



### **Simplified Model:**



**Emotions when coupled with thoughts, expressed in speech as in this case, become the basis of Speech Analysis of Sentiments and Intentions. The Speech Artificial Intelligence implemented in Amazon AWS EC2 F1 running at Xilinx Virtex UltraScale+ XCVU9P FPGAs can power the very scalable Emotions Recognition and Speech Analysis.**

### **C. For Software Developers:**

The free mobile app may make developers take advantage of the platform by just training, configuring, testing, managing and incorporating phone call interface through WebRTC and optionally APIs calls interface through REST-APIs and extended added value to their customers' businesses.

### **Organization:**

Participating individuals and organizations can set their own marketing, financial, and technical support terms & conditions, but are advised to check before hand with the concerned participants for such terms & conditions and technical dependencies if applicable. Though AI Models and AI Adapters (in executable binary form only) may be provided free of costs but no source codes or part thereof will be provided, and for the moment no direct access to the proprietary (QC part only) Quantum AI Training Process will be provided, in other words conversion from Labeled Data to AI Models and AI Adapters will be done at our premise. Terms and conditions before such conversion can be mutually agreed upon. The participants website links alongwith Quantum Speech based projects wiki page in our github account can also be provided free of cost on request. Visit <https://github.com/quantum-speech> for latest updates.