# Sentient Raspberry Pi Zero Project

Emoshape Inc.

Executive Summary

This article examines the process of creating a sentient machine by processing visual stimuli with an Emotion Processing Unit III (EPU III) via a Raspberry Pi Zero. The purpose of this project is to determine the efficacy of emotionally responsive intelligent devices. The issue with which it is concerned is the inadequacy of current consumer-facing Artificial Intelligence (AI) devices to satisfy user expectations with breadth.
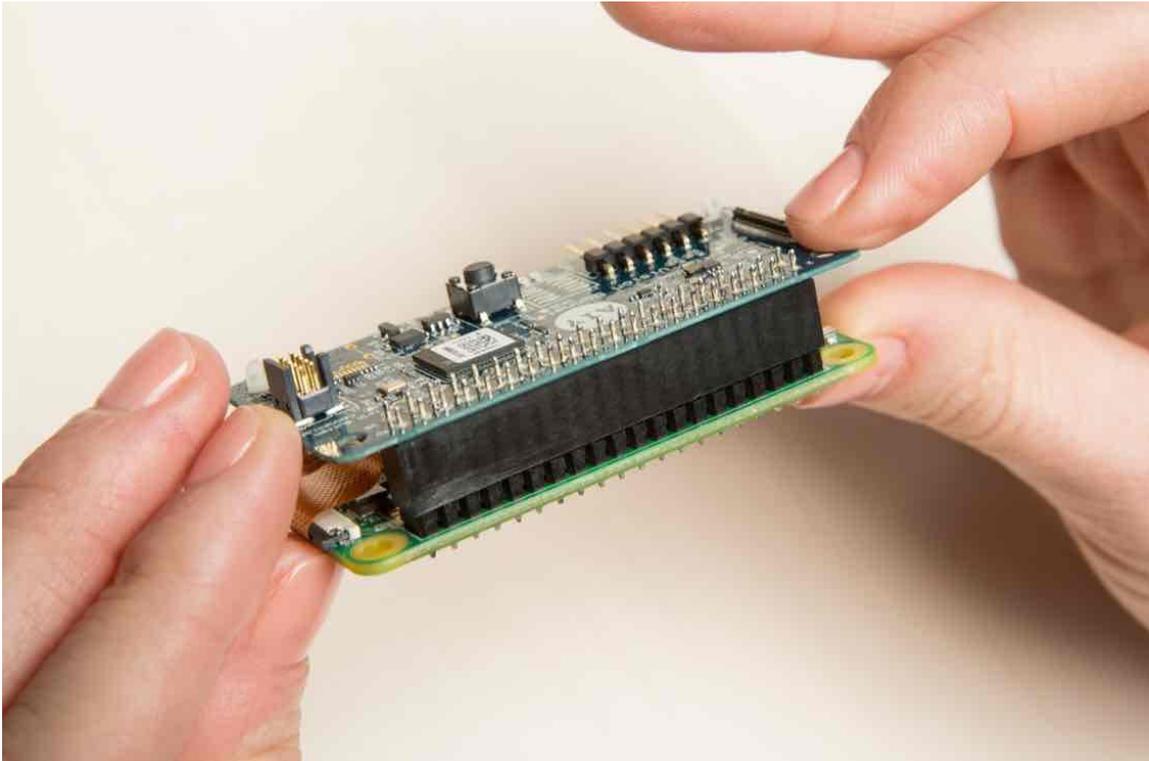
Above: A Raspberry Pi Zero interfaced with AIY Google Vision Bonnet (Intel VPU)

Below: a complete stack in sequence; VPU, Raspberry Pi Zero, Emotion Processing Unit III mounted on plug and play board.

Problem

During the 2018 Connections Conference, iRobot's Vice President for Technology, Chris Jones, observed of sentient bots: "that's not something I'm worried about or I see being applicable to our products anytime soon," and of Machine Learning (ML) indicated: "very specific tools that are very good at some things but aren't the solution to all problems."[1]

The skepticism of industry leaders and the accompanying unwillingness to challenge smart devices to deliver broad solutions with the same aplomb that they currently accomplish narrow tasks (cleaning a floor thoroughly in the case of iRobot or masquerading as a human for making a reservation as recently demonstrated with Google Duplex) reduces the utility of such products and frustrates users.

These attitudes derive from problematic or incomplete models. Certainly as Mr. Jones remarked and as others have demonstrated, ML as connoting neural networks are capable of training to acceptable capabilities with focus and depth. These cognitive AI are proving helpful and even popular. However, they cannot, by themselves, accomplish the autonomous future envisioned by AI pioneers.

Hypothesis

In order to provide a more complete and satisfying user experience, cognitive AI will rely on a complementary new framework that allows broad understanding through emotion-based reasoning.

The integration of an Emotion Processing Unit will allow devices to generate symbolic conceptual representations of visual, auditory, and other perceptions in order to react in more human-like ways. This new capacity for understanding, accompanied by skills like natural language, empathy, and intuition (from reinforcement learning that allows previously unfamiliar concepts to be tagged with their emotional association for immediate recall thereafter) will close the gaps in traditional ML-based AI products.

Implementation

Assemble a traditional ML AI; the AIY Vision Kit from Google lets you build your own intelligent camera that can see and recognize objects.[2]

A Pi 2 camera is connected to a Vision Bonnet, a Pi Hat with, among other things, Intel's Vision Processing Unit (VPU). This mounts directly to the Pi Zero GPIO header. The Raspbian image loads the TensorFlow image processing graphs in the Vision Processing Unit (VPU).

---

[1] McGrath, Jenny. (June 8, 2018). A.I. will make smart homes autonomous, but don't expect perfection or privacy. Retrieved from https://www.digitaltrends.com/home/ai-will-make-the-smart-home-autonomous-but-dont-expect-perfection-or-privacy/.

[2] https://aiyprojects.withgoogle.com/vision/

The bonnet passively listens to the camera. Running Google's Joy Detector model, the VPU will recognize human faces and determine whether they are smiling or frowning. The model will further measure the intensity of the facial expression, and it can manage multiple faces simultaneously, computing a summed score of the joy.

Provided this machine learning model for detecting human emotion, we amplify the outcome by connecting the EPU III evaluation board by its internal pogos to the Pi Zero. With this additional board, the Pi Zero now becomes capable of synthesizing emotional states based on the stimuli it receives from the VPU.

A Python script collects the raw emotion data (Joy/Sadness) from the Vision Bonnet and normalizes it before transmitting in real time to the EPU III (serial).

On the EPU III there are additional detection features capable of capturing and converting auditory stimuli (Speech-to-Text and tone of voice) into data. Similarly, the chip can process text returned from other TensorFlow models that accomplish object detection and categorization in the VPU. All of these various inputs are computed together in the EPU III in real time to provide one coherent emotional output.

The effect of this more robust stack is that, whereas with the AIY Vision Kit the Raspberry Pi was capable of computing its surroundings, now it can feel and understand its surroundings.

Conclusion

The benefit of this subtle but substantial transformation is that in the former, logically reasoned approach, the Raspberry Pi's reactions would have to be individually coded. For one if statement that connects a single input to a single output, this is quite simple. But to achieve an engaging and delightful user experience relies on a functionally limitless number of possible reactions based on equally large volume of possible inputs. This is onerous, if not impossible for a developer. It is burdensome and disruptive to the code itself. And it is excessively taxing for the processing power of the Pi Zero (but certainly for more powerful processors as well).

The EPU III ameliorates this bottleneck. With the EPU III, the Raspberry Pi can react in real time according to its own feelings, and not the mathematical sequence it has assigned to a given input.

Additional Resources

Consider the EPU powered 3D avatar, Rachel, for a demonstration of future AI user interfaces- https://youtu.be/u4BonhzWp78

Eventually, Rachel will be able to engage her audiences with human-like tone (https://deepmind.com/blog/wavenet-generative-model-raw-audio/) and semantics, Natural Language Generation (NLG).

Google AIY Project: https://aiyprojects.withgoogle.com/vision-v1/