# Project Proposal: Feel the Vibe

*Alex Strasser (arstrasse), Samuel Leong (scleong)*

## Introduction and Use Cases

Wearables are getting increasingly popular and feasible as computing power becomes smaller and more powerful. Indeed, the advent of advanced sensors and intelligence being packed into smaller packages opens up new opportunities for novel applications. In "Feel the Vibe", we aim to create a wearable that is able to display the emotion from speech in the form of an RGB LED strip that changes colors and brightness based on the detected emotion and its intensity. Applications include: being used as an assistive device for autistic or deaf people in detecting tones in someone else's voice; or as a fashion item such as a bracelet, earrings, hat, tie, belt, or shirt. It is noted that embedded deep learning devices are perfect for this application, not only due to the small form factor required for wearables, but also because local processing of audio preserves privacy.

## Dataset

We will be using the Speech Emotion Recognition dataset (actually a combination of the Ravdess, Crema, Tess and Savee datasets) available publicly from Kaggle[1][2]. The dataset supports 9 classes of emotion with 2 levels of intensity (low and strong). We will pruning emotion classes with proportionally low counts to avoid data skew; in particular, we will support the following 6 emotions (including neutral), with the following LED colors chosen based on cross-cultural psychology research[3] on the relation between emotion and color:

| **Emotion** | Anger | Happiness | Sadness | Disgust | Fear | Neutral |
|---|---|---|---|---|---|---|
| **Color** | Red | Yellow | Blue | Green | Black[4] | Off |

[1] https://www.kaggle.com/code/shivamburnwal/speech-emotion-recognition
[2] We have also found the CMU-MOSEI dataset which is a big dataset including both video and audio information, but we will probably only use it if our accuracy is undesirable
[3] https://www.psychologytoday.com/us/blog/color-psychology/202202/why-links-between-colors-and-emotions-may-be-universal
[4] Since LEDs cannot actually display “black”, they will display dark purple instead.

## Implementation

The emotion recognition algorithm will effectively be a deep learning classifier (probably a simple MLP but possibly a convolution-based network[5]) that learns and inferences from a combination of extracted audio features. These features[6] will include spectrograms, Spectral Entropy and Flux, and Mel Frequency Cepstral Coefficients (MFCC) extracted from a Short-Time Fourier Transform (STFT). In order to save power, we will need to use an event-triggered system, where a simple voice-activity detection algorithm based on linear feature prediction[7] is used to determine when a person starts and stops speaking. This will then trigger the deep-learning based emotion detection algorithm to start and stop accordingly.

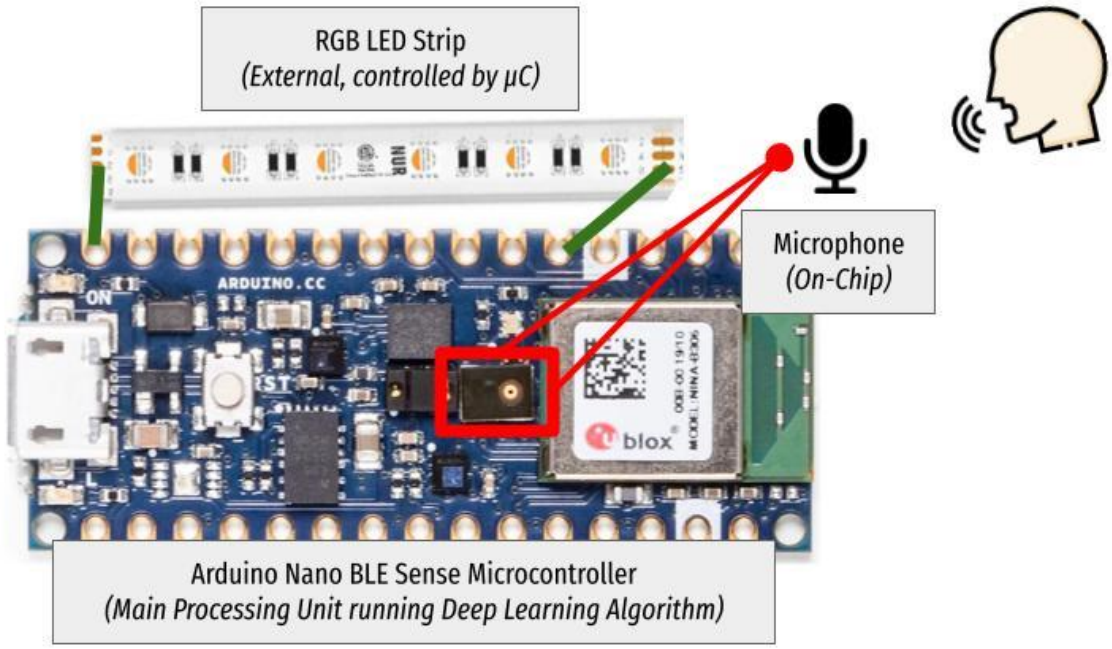


*Figure 1: High-Level Design Diagram visualizing components and connectivity.*
*Note that components are not to scale (LED strip should be significantly larger than microcontroller).*

[5] We think that a convolution-based model is overkill since the features have already been pre-extracted by convolution (FFTs)
[6] https://medium.com/heuristics/audio-signal-feature-extraction-and-clustering-935319d2225
[7] https://wiki.aalto.fi/pages/viewpage.action?pageId=151500905