Hypothesis

My hypothesis is that it will need:

- A graphical representation of a song (spectrogram)
    - The method for achieving this was to use a spectrogram. This was chosen as it contains more information than what the common *waveform* contains as it shows time, frequency, and volume.
- A human-like level of precision (50ms)
    - At first, I imagined this to be a potential variable but with the exponentially amount of required experiments & lack of time, this was set to be 50ms which seemed to be a sweet-spot for what was found in the *Format Test*.
        - Referred to as “interval”
- A subsection of time of at least 15 seconds
    - This just seemed like a good length of time since the AI would need to have a sense of the rhythm of the song to be able to find “key points”, much like a human.
        - Referred to as “backpeek”
- A decent length of time dedicated to training (128k steps)
    - While the AI is based off of Tensorflow’s image recognition base, it definitely far off of what is need to recognise a cat, for example. Thus, much time *should* be required to drop what was needed for object recognition for some more *logical* recognition.
        - Referred to as “steps”