

空気を読むAI「KI」

表情や雰囲気を読み、AIから話しかけて最適な音楽を再生するプロダクトの制作

群馬県立前橋高等学校 2年 小野瑛太 小坂橋悠斗 村田佳成琉

動機・目的

音楽を聴く方法の殆どは音楽配信サービスを利用している(*1)。音楽は人々の感情にプラスの影響を与えるが(*2)、現在の音楽配信サービスではただ音楽を流す媒体にすぎない。そこでAIを利用してその人の状況を認識し、AIから人に働きかけることで最適な音楽を再生できるシステムを開発したいと考えた。

システム概要

システムは表情・音楽・会話判定部からなる(図1)

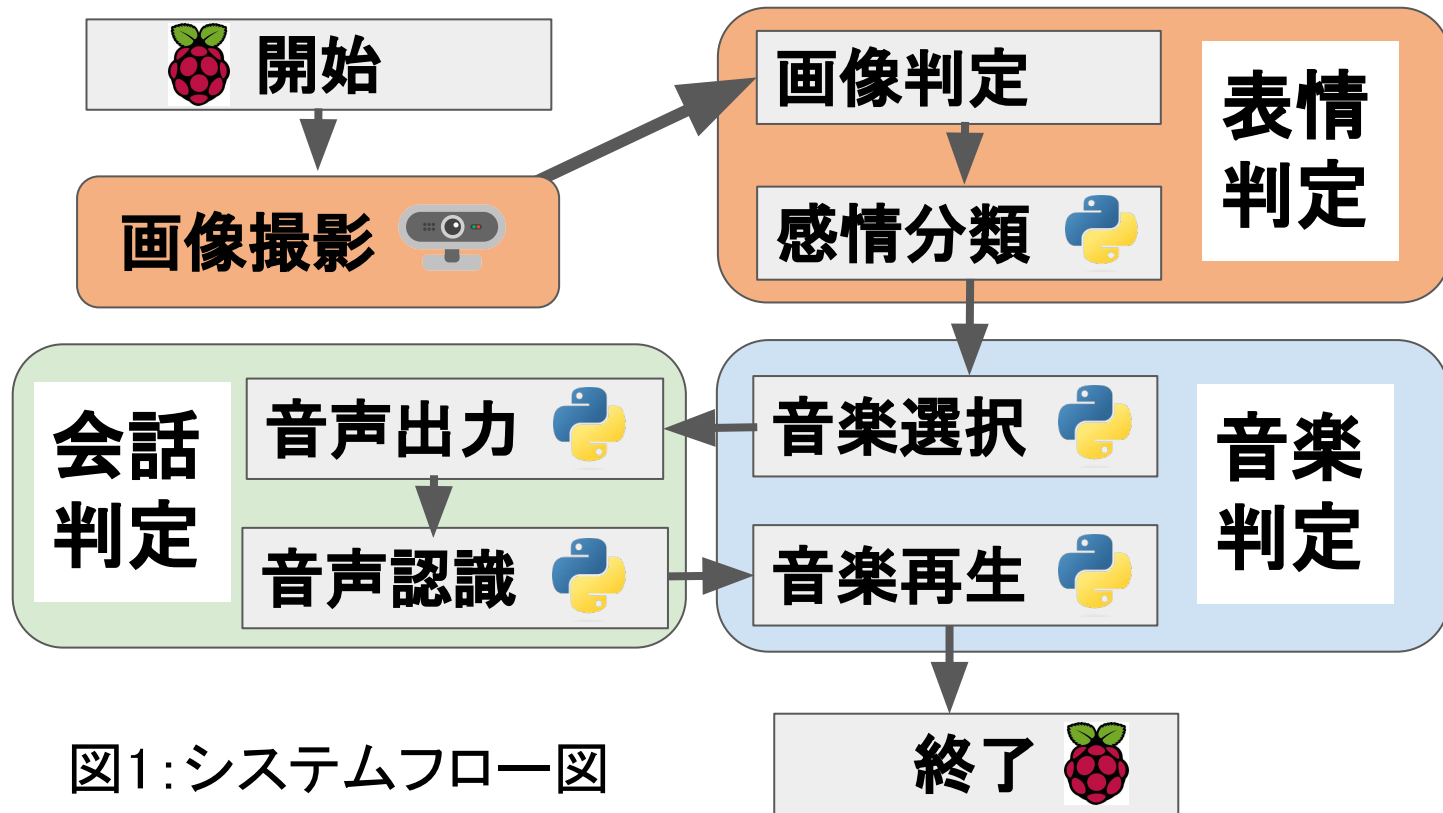
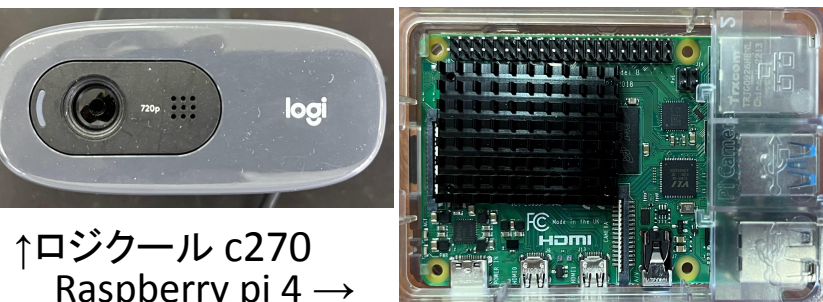


図1: システムフロー図

表情判定

RaspberryPi 4とwebカメラを用いて、Python上で表情を認識し、感情を認識する。



↑ロジクール c270
Raspberry pi 4 →

検証 PythonでdeepFaceを使用
表情認識の精度が向上。
図2のように感情を数値化し、感情のレベル分布を設定した。

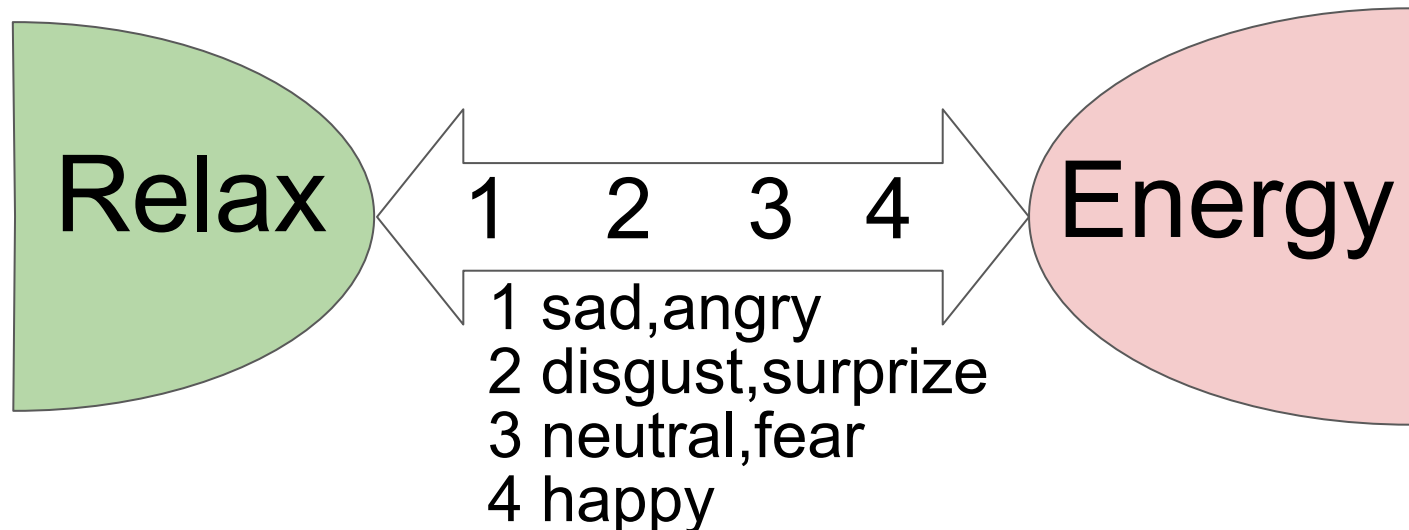
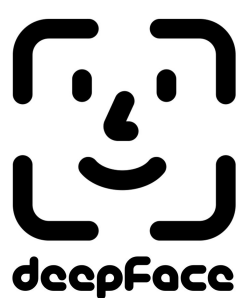


図2: 感情のレベル分布設定

音楽判定

PythonでSpotify APIを用いて以下の手順で音楽を選択する。



- ① 曲から"energy"の指標を取得
- ② "energy"の値に応じて曲を4段階に分類
- ③ 分類した曲を図2と対応させる
- ④ 感情に応じて曲をランダムに再生

例えば"happy"の場合、レベル4の曲が再生される

会話判定

デバイス側から音声で会話するように、表情に合わせた音楽をかけるかどうかを判定する。会話は感情レベルに応じたテンプレートを実装した。

検証

音声出力にOpen Jtalkを用いて音楽再生の許可(図3)をとり、SpeechRecognitionの音声認識で使用者の承諾を確認し、音楽を再生することができた。

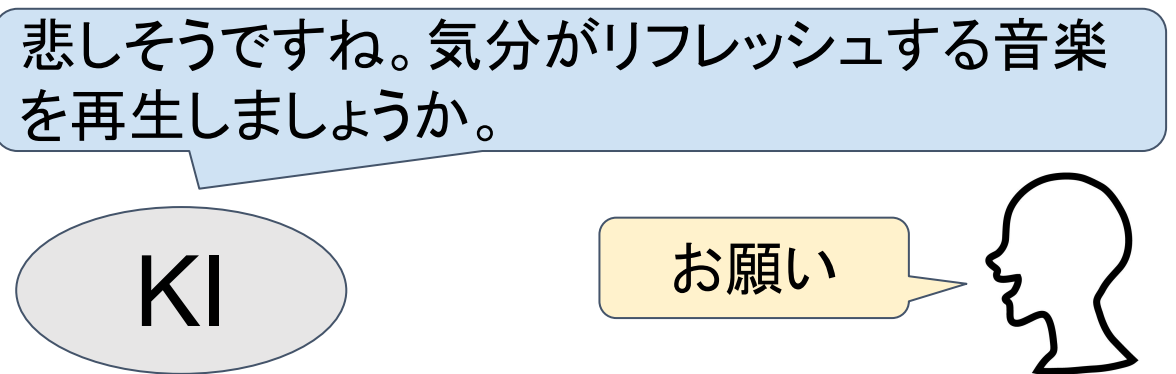


図3 会話判定部の概要

システムの課題と検証

システムの課題

表情判定において、neutralでは最適にカテゴライズされない。

検証

GPT-4V API に対して、BASE64化したwebカメラ画像と、質問” describe the person’s vibe in the image in one word”(画像に映る人の雰囲気を一言で表現してください)(*3)をインプットすることで、neutral判定されたユーザーの雰囲気をキーワード化できた。

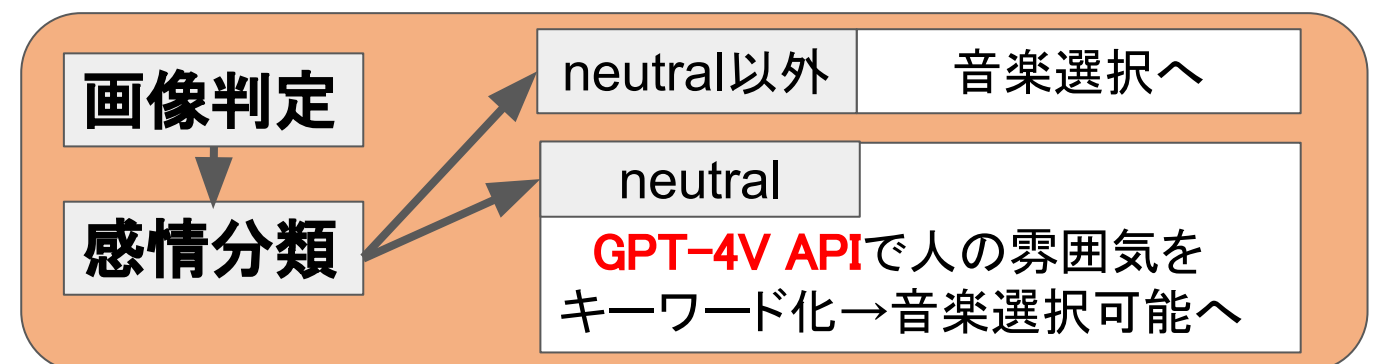


図4 表情判定部の分岐

結果

図4のように、neutralの感情判定を含め、全ての感情レベルで、デバイス側からの働きかけが可能になった。

まとめと展望

ユーザーの表情や雰囲気をAIが読取る判定部に表情と様子を判定する機能を実装することで、AIから働きかけ、感情や様子に合わせた最適な音楽を再生するシステムができた。表情認識の精度を向上させることやユーザーそれぞれの好みに適した音楽を流せるようにすること、会話のバリエーションを増やすことが課題である。また、Swiftを使用しアプリ化も行いたい。

参考文献

- *1 高校生の音楽視聴、聴く方法は「動画サイト」がトップ【LINEリサーチ調べ】(2023年7月閲覧) <https://markezine.jp/article/detail/36546>
- *2 音楽と感情についての心理学的研究(2023年7月閲覧) <https://core.ac.uk/download/pdf/229398023.pdf>
- *3 OpenAI developer platform (2023年11月閲覧) <https://platform.openai.com/docs/guides/vision>