R6 理数探究 中間報告（　　逸2　　）班

〈過去の心残り班〉

1.テーマについて

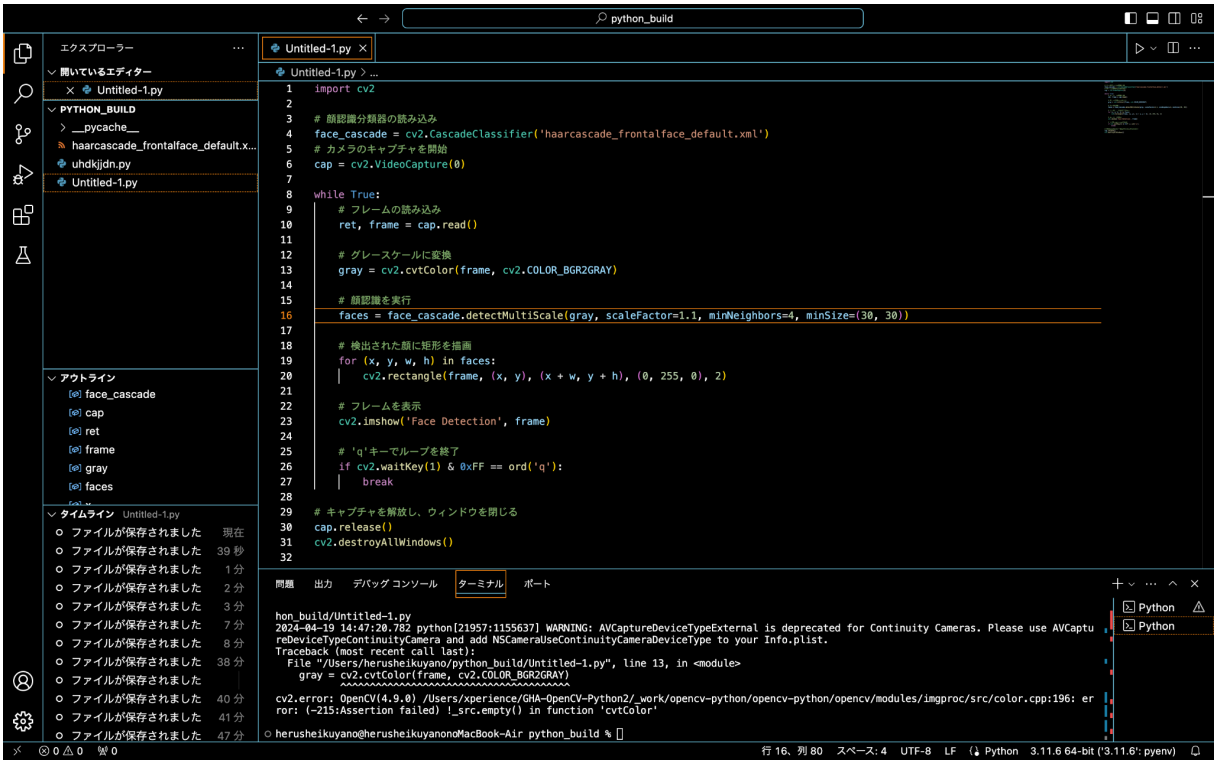
a）テーマ
顔認識システムの質の向上

b）選択した理由
スマートフォンの顔認識機能を見て、自分たちも仕組みを理解したいと思ったため。
またAIを顔認識に利用していることを知り、アルゴリズムを理解したいと思ったため。
昨年度の理数探求である顔認識システムについてやり残したことがあったため。

2.事前準備

a）使用した機材
macbookのカメラ（720p）
学校の椅子（高さ43cm）

b）使用したプログラムについて
言語:python
ライブラリ:opencv
顔認識器:
https://github.com/kipr/opencv/blob/master/data/haarcascades/haarcascade_frontalface_default.xml
を使用しました



3.探求

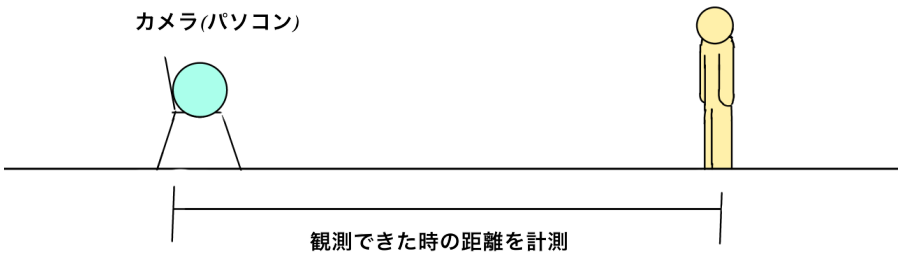
a）実験1
scaleFactor、minNeighbors、minSizeの3つの変数を用いた観測距離の調査

＊データの概要

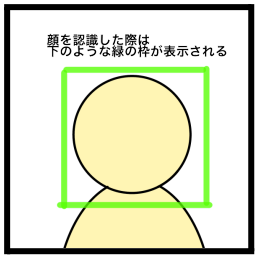
・scaleFactor　：画面の縮小量。値が大きいほど倍率も上がる。
・minNeighbors　：顔の部位の個数。値の数だけ近傍短形（顔の部位）を持つ必要がある。
・minSize　：とり得る最小のサイズ。値より小さい物体を物体として認識しなくなる。

ⅰ）仮説
scaleFactor、minNeighbors、minSizeの3つの変数にはそれぞれの最適値がある

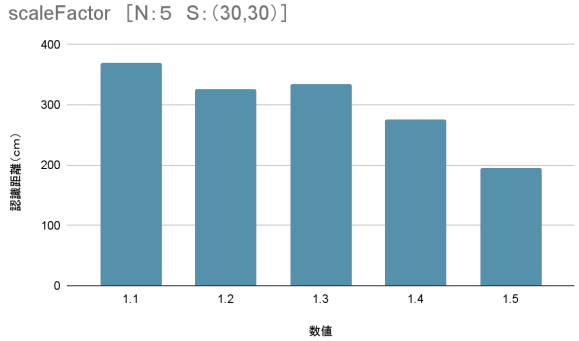
ⅱ）実験内容
上記の変数の値を一つずつ変えメガネ、マスク等を付けずに椅子の上に置いたパソコン（カメラ）の正面方向に立ち、カメラが人の顔を認識するまで近づく　その時のカメラから顔のある位置までの距離を記録した。
＊顔を認識しやすいようにスマホのライトを用いて顔の上、下側の2方向から顔　を照らした。精度が関係するデータは誤認識した回数も記録する。誤認識の回数計測　の方法は観測した距離地点から左右に1、2歩動いた際の顔以外に部分を顔と認識し　た数とした。認識したと判定するタイミングはパソコンの画面に緑の枠が途切れず　に表示されたタイミング

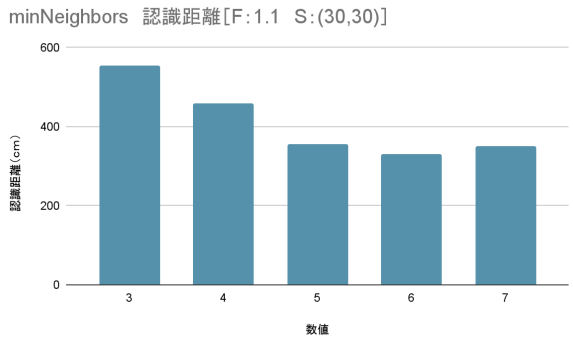


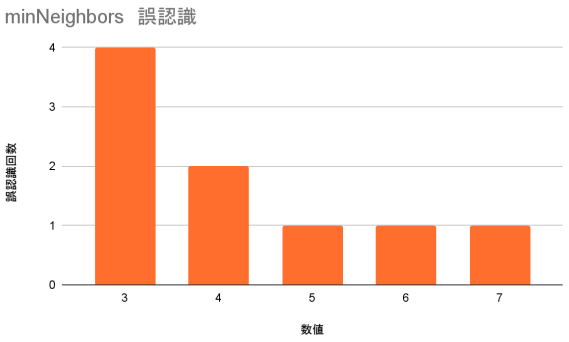
とする。

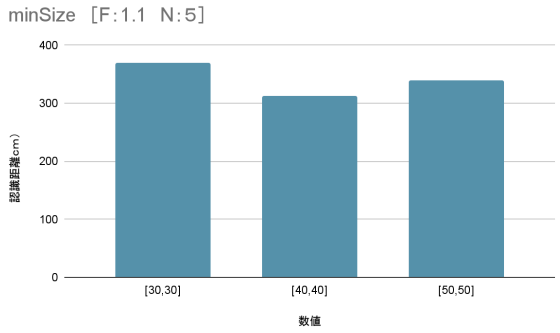


ⅲ）実験結果









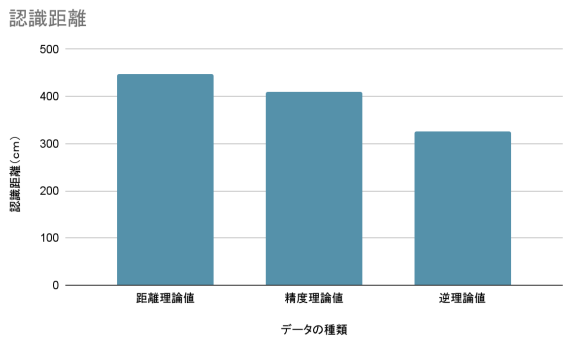
考察 ＊scaleFactor、minNeighbors、minSizeをそれぞれF、N、Sと省略する。
F:数値が基準に近いほど距離が長い。
N:数値が小さいほど認識する距離が長いが、小さいほど精度が落ちる。
S:相関は見られないが、基準の値が最も良い結果となっている。

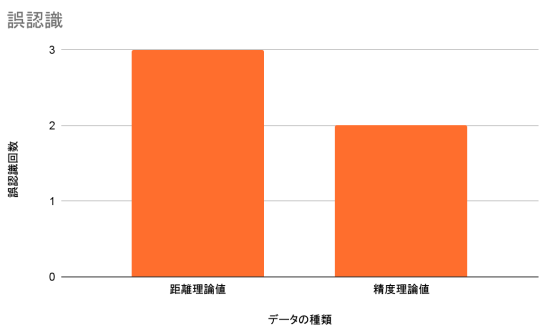
b）実験２
実験１の結果を用いた観測距離の向上

ⅰ）仮説
3つの変数の一番いい値を用いればより長い距離で観測できる

ⅱ）実験内容
実験１の結果を用いて3つの変数の最もよい組み合わせ、即ち理論値の組み合わせを精度重視と距離重視の2種類を用意する。また最も悪い組み合わせ、即ち逆　　理論値の組み合わせの計3つを実験１と同じ手順で実験を行う。

ⅲ）実験結果
・距離最適値
（F:1.1、N:3、S:[30,30]）
・精度最適値
（F:1.1、N:4、S:[30,30]）
・逆最適値
（F:1.5、N:6、S:[40,40]）





4.考察

認識する距離を伸ばすためには認識の難易度を落とさなければならず、精度と距離どちら　かを優先する必要がある。

使用用途によって変えると良いだろう。

5.今後について

VOICEROID班に合流し探求を進めていく。

時間に余裕があれば、顔に装飾を加えたり、2次元画像を用いた実験を行っていきたい。

<VOICEROID班>

1 テーマについて

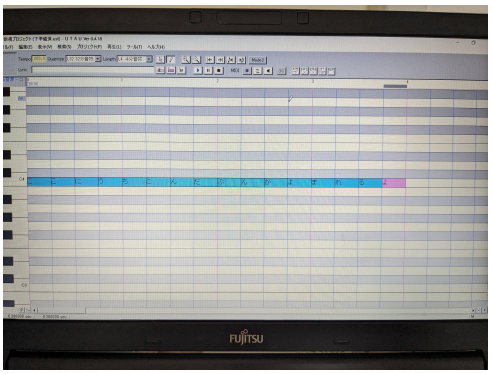
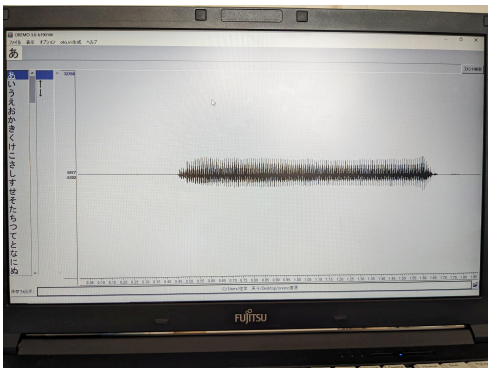
a)テーマ
自分達の声を元にしたボイスロイドを制作し評価する

b)理由
発案者がボーカロイドが好きで自分達の声で作ってみてどんな差がでるのか気になったか ら。

2 使用するアプリ
①UTAU
②OREMO
・静かな空間（今回は放送室）で録音専用アプリのOREMOを用いて録音をする
・録音したデータを音声編集アプリであるUTAUで使える形に編集する

3 探求したこと

編集した音声データをUTAUで切り貼りし、1つ目のボイスロイドを作った。

左:UTAUで文章を打ち込んでいる様子
右:OREMOで録音している様子（録音した声の波形が表示されている）

4 結果と考察

1つ目につくったボイスロイドでは音が途切れ途切れになってしまっていたり、音の高さや大きさにばらつきがあったり、そもそも聞き取れないような音が含まれていたりした。
そのため、もっとはっきりとした声でかつ同じくらいの声量と高さで録音し直し、ボイスロイドを作ることで今回作った1つ目よりももっと聞き取りやすいボイスロイドに、なるのではと考えられる。
また、途切れ途切れになってしまう部分は音単体を編集する段階で適切なデータに編集できてなく無駄な余白が生まれてしまっていると考えられるため、次のボイスロイドはもっと編集する時間をとり、文章にしたときの余白が少なくなるにはどのような編集をするのがよいかというのも探していくべきだと考えられる。

5 今後の予定

・UTAUを用いてボイスロイドをもう2つ制作し、それぞれ作成したボイスロイドの性能にどのような差が発生しているかを調べる。
・考察で出てきた改善案を実践し、差が出てくるかを調べる。
・上の心残り班の発表を、作成したボイスロイドで発表する。
・別の制作方法であるRVCを使い、ボイスロイドを制作する。

＊RVC(Retrieval-based-Voice-Conversion)について
　中国で開発された、AIを使用した音声の機械学習、音声変換が出来るソフト
・VC Client(Voice Changer Client)
　RVCで学習させたデータを使ってボイスチェンジャーの役割を果たすソフト
UTAUでは一音一音あいうえお順に学習していくのに対し、RVCでは10分以上の音声データをAIに読み込ませて学習する。そのため録音環境を同じにすることは出来るが同じ音声データで学習結果を比較することは難しい。
・RVCの学習データを作成しVCCを使ってUTAUで作ったボイスロイドとの音質や録音した人の声との一致度合いを比べる。