

声から顔を出力するAI



# きっかけ

声と顔には何かしらの関係  
があるのでは？



# きっかけ

## 【第2章 声におけるノンバーバルの要素の形成要因】

[https://waseda.repo.nii.ac.jp/record/7169/files/Honbun-5363\\_03.pdf](https://waseda.repo.nii.ac.jp/record/7169/files/Honbun-5363_03.pdf) (伊東 裕司 1995)

→この論文から、  
声には、骨格や幼少期の環境、  
社会上の役割などが相関している。

## 【顔と声の関連性の判断： 人物の同一性について】

[https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/download.php/AN00150430-00000098-0123.pdf?file\\_id=37337](https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/download.php/AN00150430-00000098-0123.pdf?file_id=37337) (鈴木静 2010)

→この論文から、  
顔（骨格）と声には具体性は  
わかっていないけど何かしらの  
法則性がある！

# きっかけ



高校生の私達に  
解明は難しい！！

声から顔を出力する  
AIをつくりたい！！

防犯カメラに入ってる声  
とかから犯人の顔が判明  
できたら、もっと解決で  
きる事件が増える

顔認証や指紋認証だけ  
じゃなくて声でも  
スマホを開けないかな

顔から声が判別できたら  
自分にあった歌い方とか  
英語の読み方とかが分か  
るんじゃ？



# 仮説

声と顔の因果関係なしに

声から人の顔を出力するAIは作れる。

# 実験

## ①素材を集める

・色んな人の顔（色んな角度から）  
と声（1分半程度）を集める

→出来れば色んな人から

- ・アジア人
- ・10～40代
- ・整形等をしていない



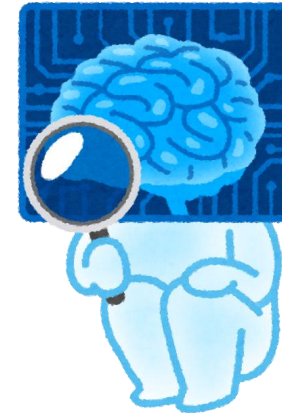
# 実験

ChatGPTからの声から顔を生成するAI  
へのアイデア ✨

## ②関数を作る

数値化  
(embedding化)

関数  
(MLP)



### ● Speech2Face (MIT, 2019)

「音声から顔を推定する」有名な研究。音声特徴量 (Spectrogram) をCNNで処理し、顔のエンベディングを生成 → 顔画像に変換。

・論文: <https://arxiv.org/abs/1905.04400>

・コードは非公開ですが、論文にアルゴリズムの構成あり。

### ● Voice2Face (2022)

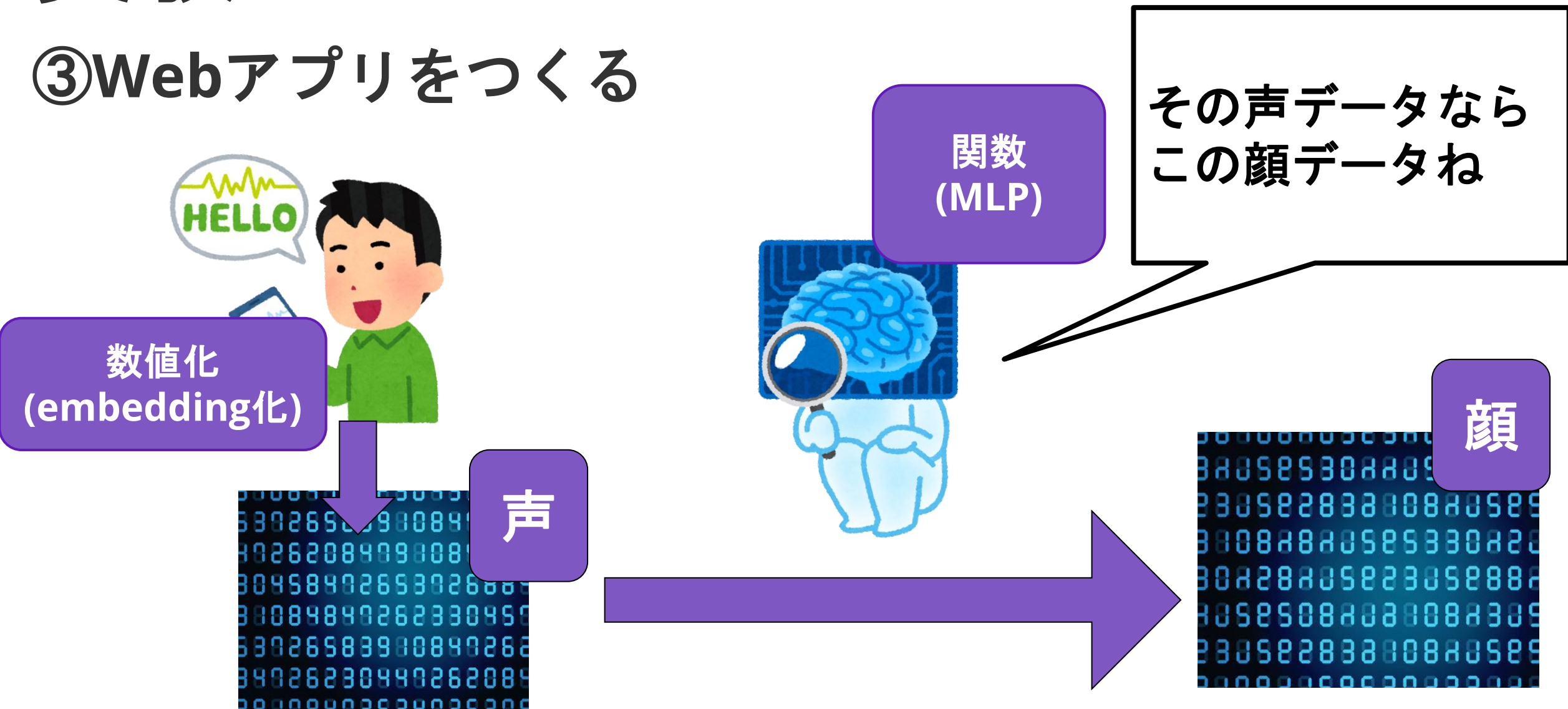
より新しい手法で、音声から顔を生成するモデル。StyleGAN2ベースで構築。

・特徴: 高解像度、表情まである程度推定できる

・音声 → エンコーダ → 潜在空間 (StyleGAN)

# 実験

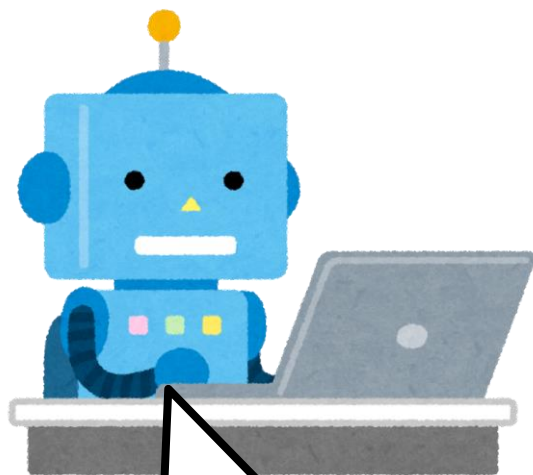
## ③Webアプリをつくる



# 実験

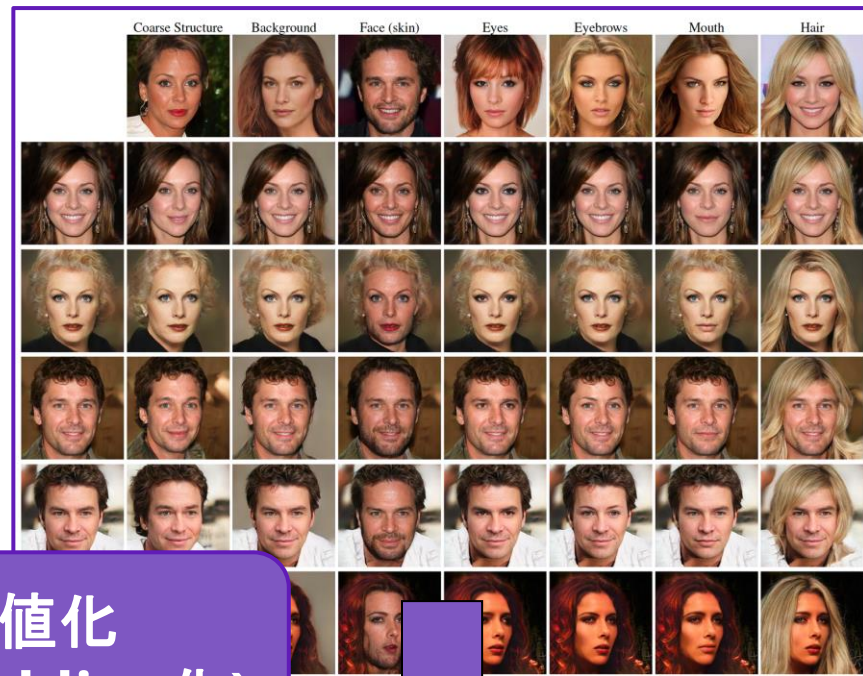
## ③Webアプリをつくる

顔



この中で一番  
近いのは...

styleGAN2



数値化  
(embedding化)





# 結果・考察

- ・ 外国人の顔ばかり出るようになってしまった。  
→ styleGAN2で画像を生成する際、**公式の既に学習されたモデル**を使用したため。
- 自分たちの集めた素材を組み込めばアジア人風の顔画像が出てくるかも。
- ・ 集めたデータ(48名)内であれば特定できた。  
→ **国民のデータ**を用いれば、近い事ができる可能性。

# 結論

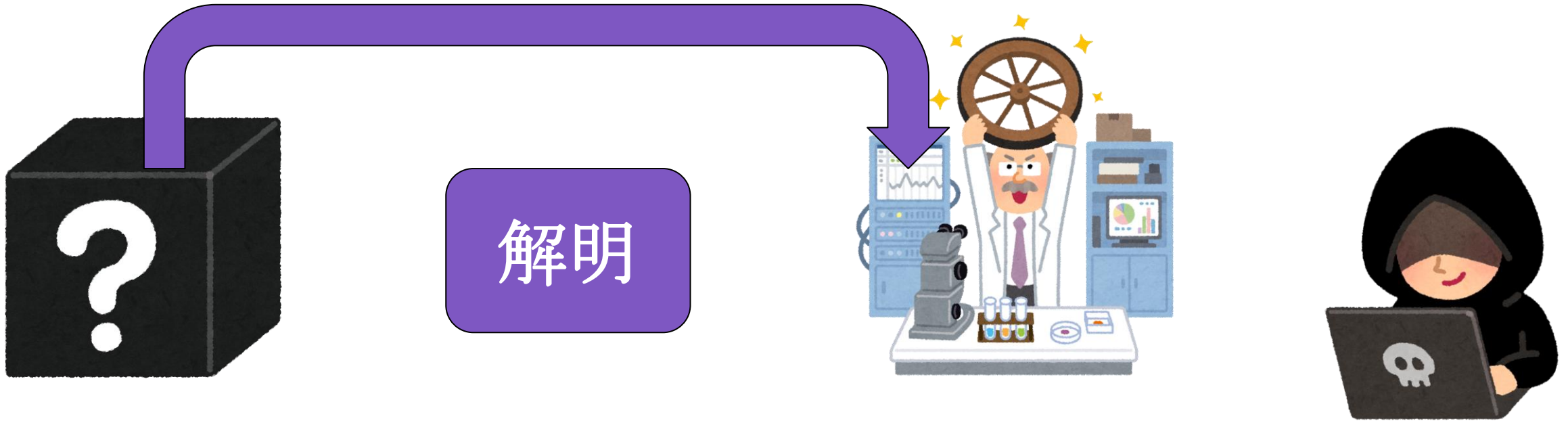
- ・ 授業内の時間では声から顔を出力するAIを作ることは難しかった。

→styleGAN2を一から学習させた場合、可能性はあるが、コストが非常に高い。



# 今後の展望

- ・ 顔と声の相関関係を使わずにAIを作れる!!



- ・ お金を払えば誰でもAIを作れることの危険性

参考

ご清聴  
ありがとうございます  
ございました