

声真似の相違～音声認識による解析～

岡山県立玉島高等学校 理数科2年生 蟹江一聖 木下和晃 栗山千輝 忠政志門 花田大樹
指導者 尾崎未登利

1. 目的

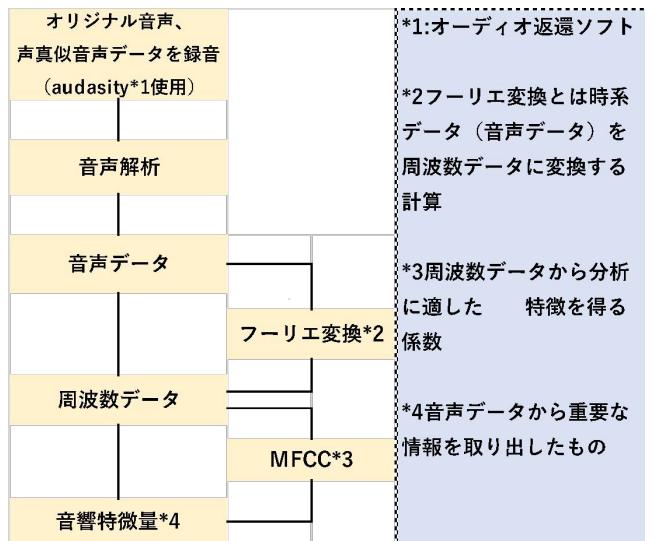
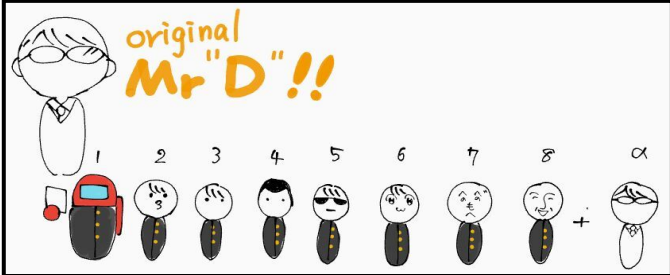
- I. テレビなどでよく見る声真似はなぜ似ているのか？
- II. 「声真似が似ている」はどうやって判断しているのか

2. 仮説

人間の**感覚的**な相似の判断結果と**物理的**に分析した相似の判断結果を比べると似ていないという結果になると考えた。

3. 研究方法

I. 音声認識流れ



*1:オーディオ返還ソフト

*2フーリエ変換とは時系データ(音声データ)を周波数データに変換する計算

*3周波数データから分析に適した特徴を得る係数

*4音声データから重要な情報を取り出したもの

II 比較方法

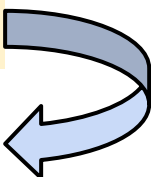
MFCCの次元ごとに差を出し、その合計を比較する。



図 音声録音している写真

9人分の音声データとオリジナルの音声聞いてもらう。

似てる、似てないの判断を点数化して順位にする



4. 結果

MFCC、アンケート、音声波形を示す。

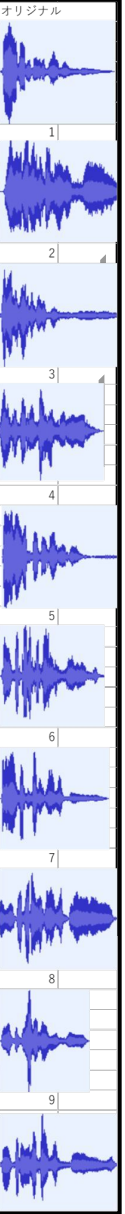
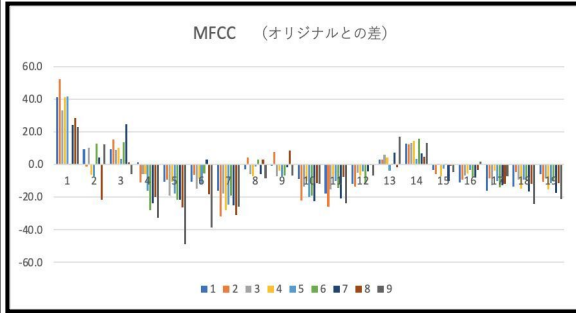


表 MFCCとアンケート結果のまとめ

番号	MFCC 順位	アンケート 順位	MFCC (19成分和)
1	1	2	-52.6
2	2	1	-65.3
3	3	6	-70.5
4	4	2	-71.8
5	5	8	-112.9
6	6	5	-121.7
7	7	9	-121.8
8	8	4	-131.5
9	9	7	-198.8

5. 考察

根拠となる音声データ2, 3, 4, 5, 7似ている似ていないの判断基準をどちらも音色とイントネーションと仮定すると、人間は音色より**イントネーション**を判断基準としていると考えられる

6. 結論

声真似が似ているか似ていないかの判断は、人間の感覚とMFCCに差はあると言える。

7. 今後の課題

本研究を踏まえて、人間の感覚とMFCCの差をより明らかにするために、音声データを増やして研究する。次元の数をオリジナルと同じにしたらどうい変化をするのか。結果を目視ではなく、すべて機械学習を通して行う。

8. 引用文献・参考文献

[アーカイブ] [Pythonサンプルスクリプト] librosaによる音声データからの特徴量抽出 (MFCC)[[音響特徴量「メルスペクトル」と「MFCC(メル周波数ケプストラム係数)」の解説と実例紹介Python]による音声認識 MFCC編
声真似は本当に似ているのか? ~音声分析による解明~ スペクトル解析補足記事