人工知能・音声処理実験2020
口頭試問

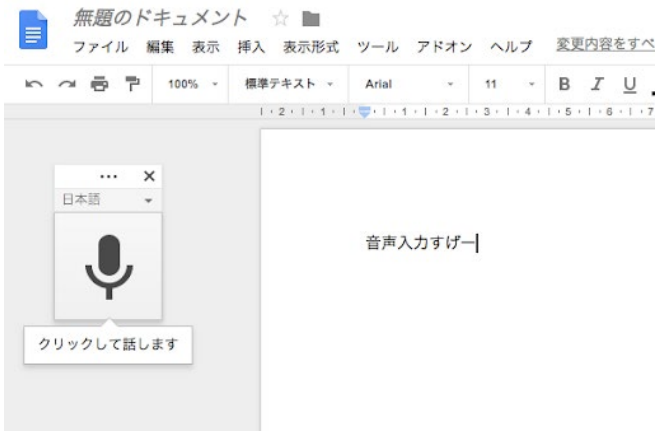
音声認識で文字列の生成をする

発表者　　:今田将也
学籍番号　　:09430509
発表日　　:２０２０年１１月２４日

概要

- “あ”，“い”，“う”，“え”，“お”の5音を認識できる音声認識器を作成する.
- なお，認識した音声はリアルタイムで画面上に表示する.

（簡単なイメージとしては，下図のようなGoogleドキュメントでの音声認識　の　5音のみ）



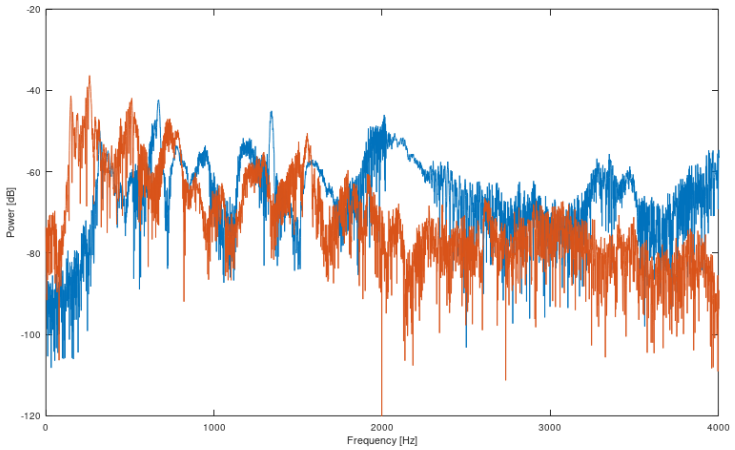
本実験で使用する単語について

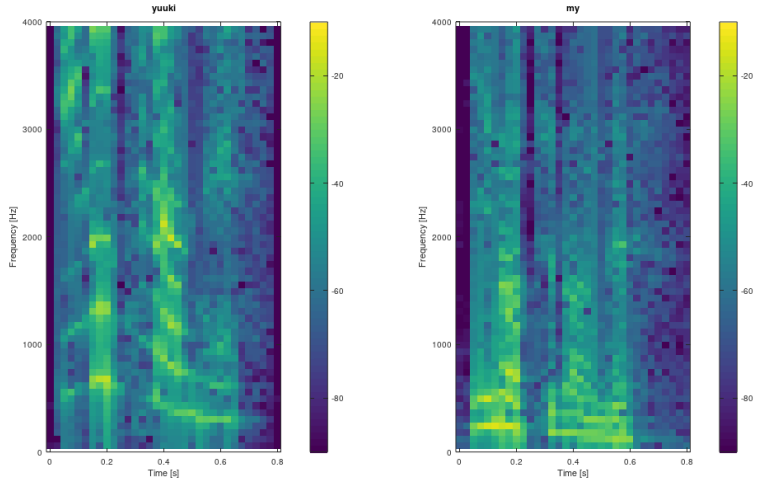
対数パワースペクトル

- ある音声波形の周波数ごとのパワーの大きさを表したもの

パワースペクトログラム

- 周波数分析を時間的に連続して行い、色によって強さを表すことで、強さ、周波数、時間の3次元的に表示をしたもの
- 音の時間的な変化、音色、高さ、大きさを同時に読み取ることができる



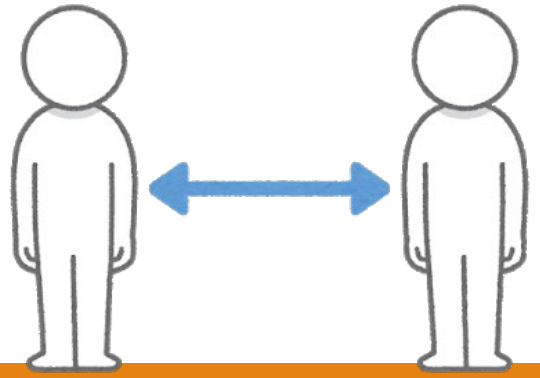


音声の識別方法

2つの音声がどれくらい似ているかをどのように判別するのか？

- 入力音声と参照音声間でパワースペクトルを出し，それぞれの周波数成分についての誤差の二乗和の根を計算する

$$D(x,y) = \sqrt{\sum_{n=1}^{N} (x_n - y_n)^2}$$



実施した実験

1. あ，い，う，え，おの5音を2個ずつ録音
2. １音ごとに参照パターンを生成し，５音を認識できる音声認識器を作成
3. １秒ごとに音声を，プログラムを止めるまで入力
4. 即座に判定結果を画面上に出力
5. どれくらい正しく認識しているかを見る

実験結果

入力していった音声
- あえいうえおあお

出力された文字列
- あえいうえおあお

```
>> ASR

say
RESULT = A
say
RESULT = AE
say
RESULT = AEI
say
RESULT = AEIU
say
RESULT = AEIUE
say
RESULT = AEIUEO
say
RESULT = AEIUEOA
say
RESULT = AEIUEOAO
```

実験結果

入力していった音声
- あういえおいうあおあおいえ

出力された文字列
- あううえおいうあうあおいえ

```
>> ASR

say
RESULT = A
say
RESULT = AU
say
RESULT = AUU
say
RESULT = AUUE
say
RESULT = AUUEO
say
RESULT = AUUEOI
say
RESULT = AUUEOIU
say
RESULT = AUUEOIUA
say
RESULT = AUUEOIUAU
say
RESULT = AUUEOIUAUA
say
RESULT = AUUEOIUAUAO
say
RESULT = AUUEOIUAUAOI
say
RESULT = AUUEOIUAUAOIE
```

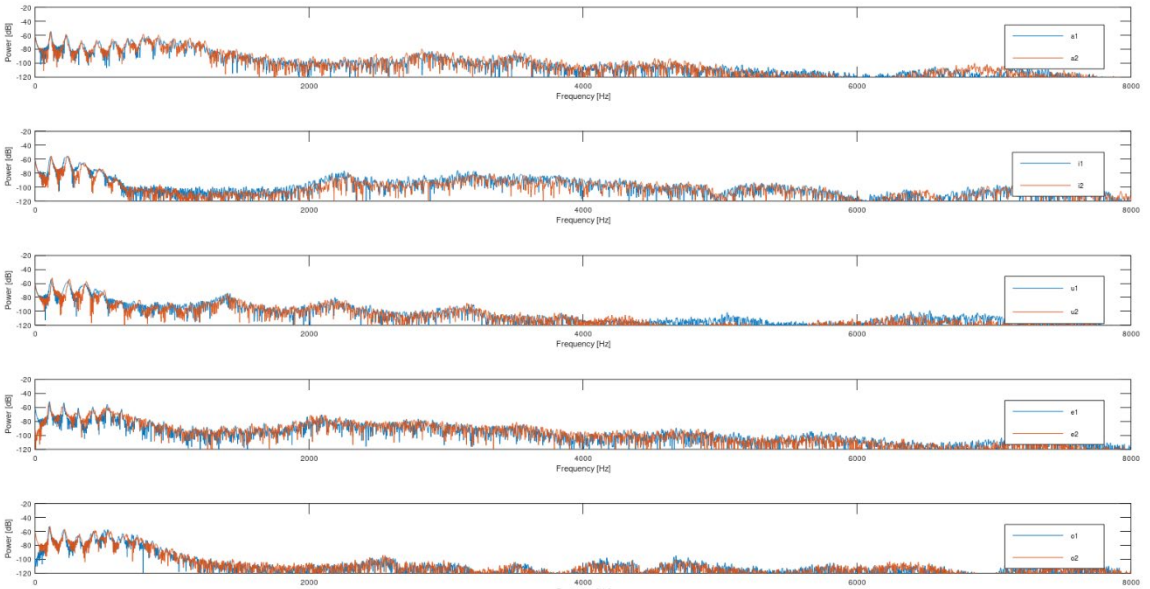
結果に対する考察

【考察】

◆い・お　が　う　として認識されてしまっていることについて

- “い”の参照パターンと“う”の参照パターンの2000hz以降の特徴が似通っている.
- “お”の参照パターンと“う”の参照パターンの500hz以下のパワーの山付近の特徴が似通っている

どちらかの判別がつかなくなり，どちらともとれそうな“う”が結果として判別された可能性



上から，“あ”，“い”，“う”，“え”，“お”のそれぞれ2つの対数パワースペクトルの図

結果に対する議論

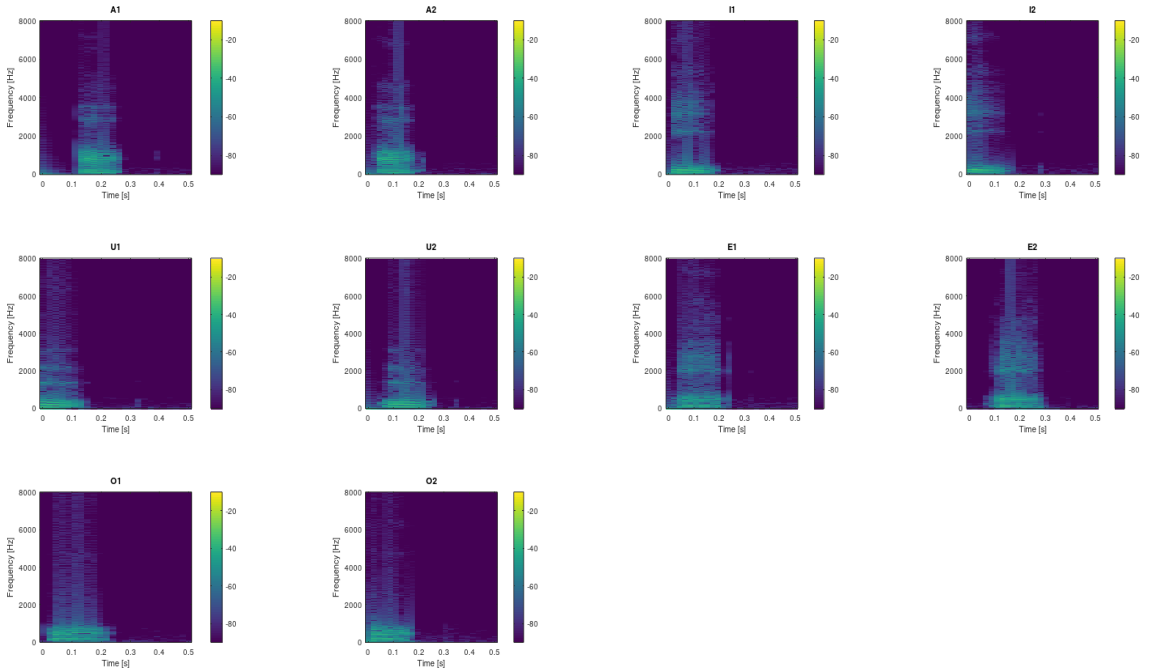
【議論】

◆精度をあげるにはどうすれば良いのか

- 5音のそれぞれのスペクトログラムを見て，時間を経過させても継続的に特徴が現れていそうな周波数帯で距離を計算させることができれば音の識別も向上すると推測

◆1秒ごとにタイミングが合わないと認識されない

- 音と音の途切れ部分を認識させる手法が必要になる（今回は検討していないが・・・）



左上から，“あ”，“い”，“う”，“え”，“お”のそれぞれ2つずつのスペクトログラム

まとめ

◆実験の前提となる，パワースペクトル，スペクトログラム，２音声の認識手法について説明をした

◆“あ”，“い”，“う”，“え”，“お”について，精度はそんなに高くないレベルでの認識をしてくれる音声認識器を作成した

◆入力するまでに認識した音声を全て画面上に表示させた

【感想】

継続的に，時間を区切らなくても音を認識してリアルタイムに文字起こしをするソフトがあったが（Googleドキュメントの文字認識）どうやって日本語や英語まで識別しているのか興味が湧いた