

# 音声処理実験 作業日報 第3回 SP-1

---

学生番号: 09430509

氏名: 今田将也

提出日: 2020年11月10日

# SP-1の概要

---

## 1. 問題1

1. MATLABで音や音声を扱うための基礎知識を4つの観点に整理して説明をした

## 2. 問題2

1. MATLABのソースコードを示しながら、音の生成からDFT(FFT)を用いたパワースペクトルの表示までの手順を説明した
2. いくつかのパワースペクトルを図示して考察をした

## 3. 問題4

- じゃんけんの3手（「グー」「チョキ」「パー」）の単語発声を音声ファイルとしてMATLABを用いて録音した

# 問題1

---

MATLAB/Octave上で音や音声を扱うための基礎知識を，簡潔に説明せよ．

# 問題1の小問題

---

1. MATLAB/Octaveにおける音のI/O処理について、以下の4つの観点に整理して簡潔に説明せよ
  1. 音の録音（ToDo: 第4回(4A節)で実施）
  2. 音の再生
  3. 音の音声ファイルへの保存
  4. 音の音声ファイルからの読み出し
2. デモ音（あるいは、自身の録音音声）を利用して、その音声波形を時間波形として表示せよ。なお、以下3つの情報を明記すること。
  1. 音のサンプリング周波数（単位は Hz か kHz）
  2. 音の長さ（単位は s か ms ※秒単位かミリ秒単位のどちらかということ）
  3. 発話内容（例えば、「漢字と読み仮名」あるいは「ひらがな」などのいずれかで書く。）

# 1-1. 音の録音・再生

---

音の録音(4Aの内容)

1. Audiorecorder関数で録音用のオブジェクトを作成
2. Recordblocking関数を用いて録音
3. getaudiodata関数を使い, 音データを取得

音の再生

Sound関数を用いることで可能.

1. 第1引数:再生するデータの変数
2. 第2引数:再生するときのサンプリングレート

`sound(y, Fs)`

# 1-1. 音声ファイルへの保存・読み出し

---

## 音声ファイルへの保存

Audiowrite関数を利用すると、音声データを音声ファイルに書き出せる。

- 第1引数：ファイル名
- 第2引数：指定した変数のデータ
- 第3引数：サンプリング周波数をHz単位で指定

## 音声ファイルの読み出し

```
[y, Fs] = audioread('example_1.wav');
```

MATLAB/Octaveでは、コマンド一つで実行できる。

- audioread第1引数：読み込みたい音声ファイル名
- 1つ目の戻り値yには、第1引数で指定したファイル内の音声データが、ベクトルとして読み込まれる。
- 2つ目の戻り値Fsには、同ファイルが保存しているサンプリング周波数が、スカラーとして読み込まれる。

# 1-2. デモ音の音声波形を時間波形として表示

---

サンプリング周波数

- 24000Hz

音の長さ

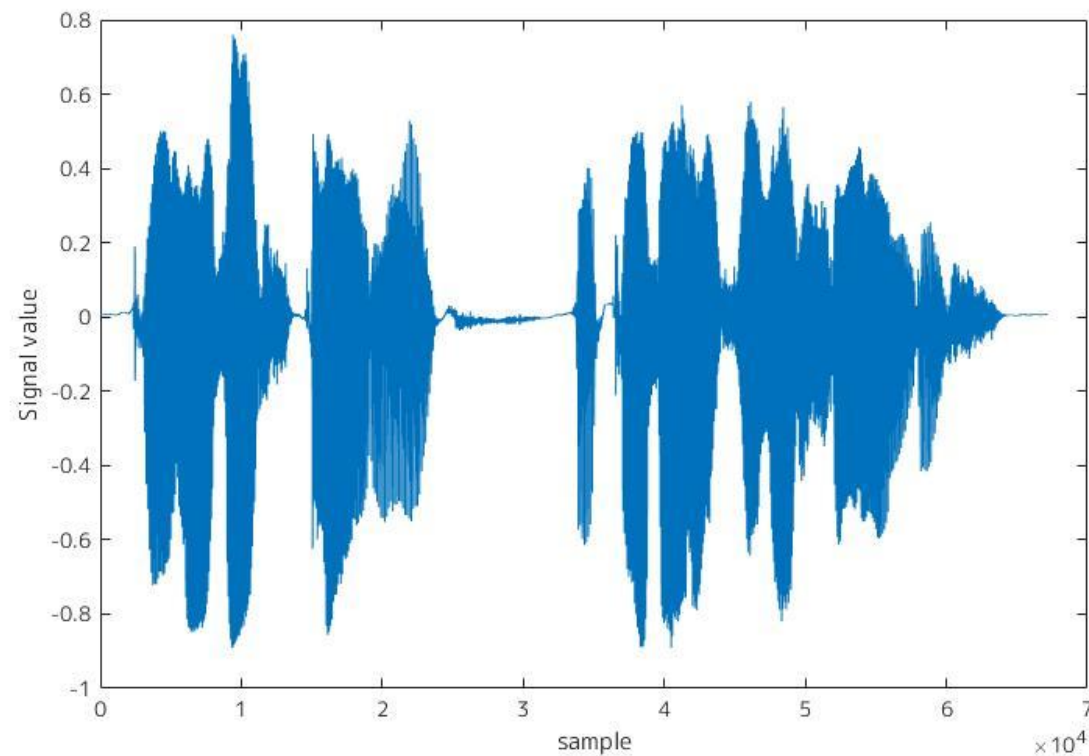
- 2.7994秒

発話内容

- このばすていはおかだいひがしもんです

サンプル数

- 67185



# 問題1のまとめ

---

## 問題

MATLAB/Octave上で音や音声を扱うための基礎知識を，簡潔に説明せよ.

## まとめ

サンプル音声を用いて，音の再生をした

実際に録音を行い，ファイルへの保存，読み出しまでの一連の流れを示した.



# 問題2

---

純音（正弦波）のパワースペクトルを観察し，パワースペクトルから得られる情報と聴感的な印象の関係について考察せよ．

# 問題2の小問題

---

1. MATLABのソースコードを示しながら、純音の生成から、DFT(FFT)を用いたパワースペクトルの表示までの手順を説明しなさい。
2. 純音の時間波形とパワースペクトルを図示しなさい。作成に関わるパラメータを明記すること。
3. MATLABのソースコードを示しながら、DFT(FFT)を活用して、実世界尺度を考慮したパワースペクトルを表示するための手順を説明しなさい。
4. 波数を変えた純音のパワースペクトルをいくつか並べて図示して、以下の観点から考察しなさい。
  1. 正弦波信号の周波数と図示された正弦波のパワースペクトルの形状にはどのような関係があるか？
  2. 正弦波信号の周波数と聴感的な印象（「高い音」や「低い音」）にはどのような関係があるか？

## 2-1. 純音の生成から, DFT (FFT)を用いたパワースペクトルの表示までの手順(1/2)

### 正弦波信号の生成

- アナログ信号の信号長[sec]をデジタル信号の点数に換算する. ここでは9600個の点
- サンプリング周波数 16,000 Hz で, 0.6秒の音を表現するためには, 9,600個の数字列が必要

### sin関数を使い, $x(t)=A\sin(2\pi ft)$ を満たす正弦波を生成

- 1:9600で1から9600までの数字列 (9600個の点) を作る
- 時刻が0から始まることを考慮し0から9599までの数字列とする
- サンプリング周波数 16,000 Hzでは, 点間は秒間にして1/16000秒に相当するため, 16000で割る
- sin関数の最後のカッコの後に, シングルクオートをつけて転置し, 音データと縦方向にベクトルの向きを揃える

```
signal_length_sec = 0.6;  
sampling_rate = 16000;
```

```
signal_length_pt = signal_length_sec * sampling_rate  
t = ((1:signal_length_pt) - 1) / sampling_rate;
```

```
A = 0.2;      % Amplitude  
f = 10;       % Frequency [Hz]  
t = ((1:9600) - 1) / 16000;  
x = A * sin(2 * pi * f * t)';
```

## 2-1. 純音の生成から, DFT (FFT)を用いたパワースペクトルの表示までの手順(2/2)

### 時間波形として出力

- Plot関数の第1引数にt, 第2引数にxを入れる.

```
A = 0.2;      % Amplitude
f = 10;       % Frequency [Hz]
t = ((1:9600) - 1) / 16000;
x = A * sin(2 * pi * f * t)';

plot(t, x)|
```

### DFTでのパワースペクトル出力

- MATLABに標準搭載されているFFT関数を使うことでDFTを実装可能
- 信号をfft関数に適用し, その配列数(length)で割る
- 結果は複素数のベクトルであるため, 絶対値の二乗を行う
- 上記の結果をplot関数で表示する

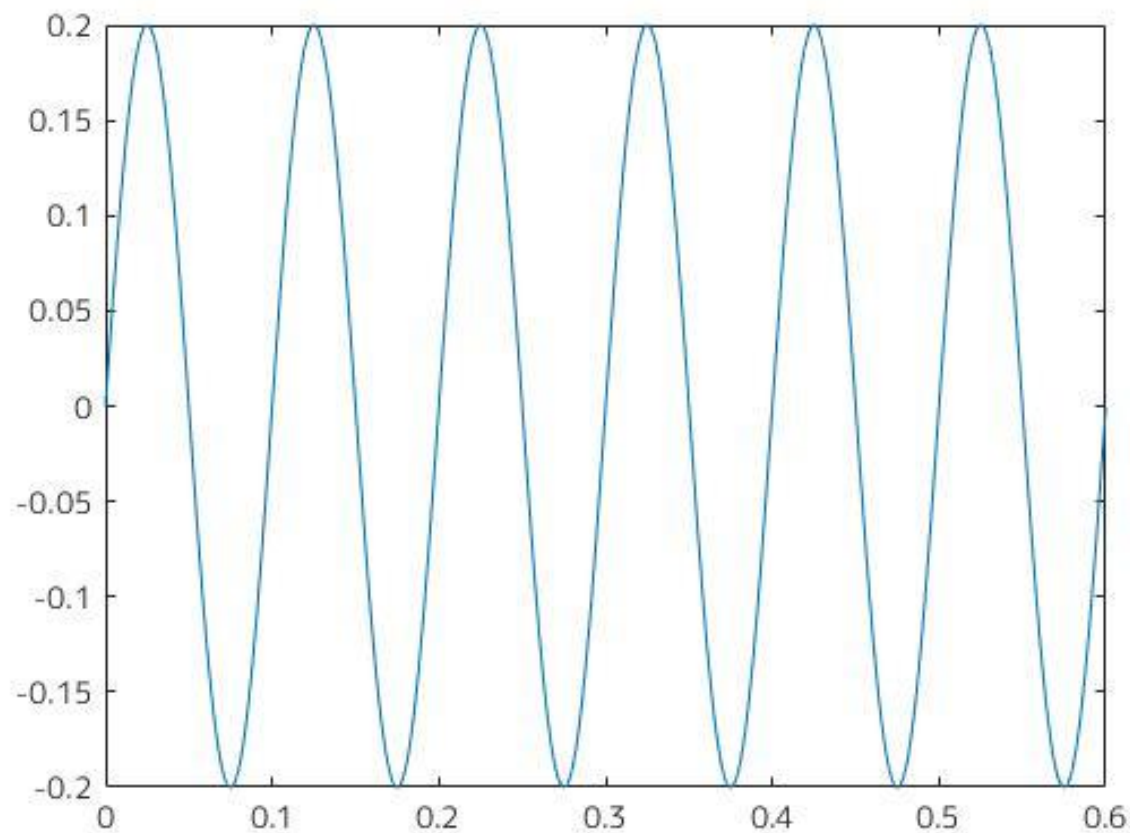
```
%DFT
X = fft(x) / length(x);
XPow = abs(X) .^ 2;
plot(XPow, 'o-')
xlim([1 10]) %plot array index from 1 to 10
```

## 2-2. 純音の時間波形と設定パラメータ

### パラメータ設定

- 秒数 : 0.6sec
- サンプリングレート : 16000Hz
- 振幅A : 2
- 振動数 : 10hz

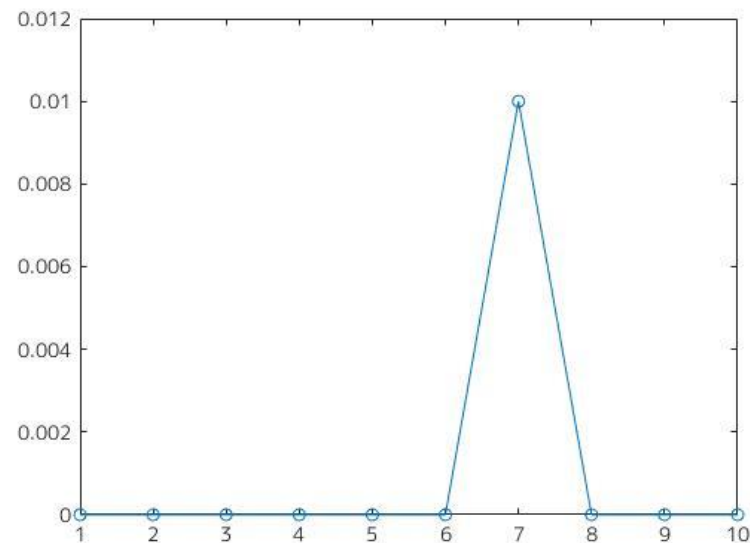
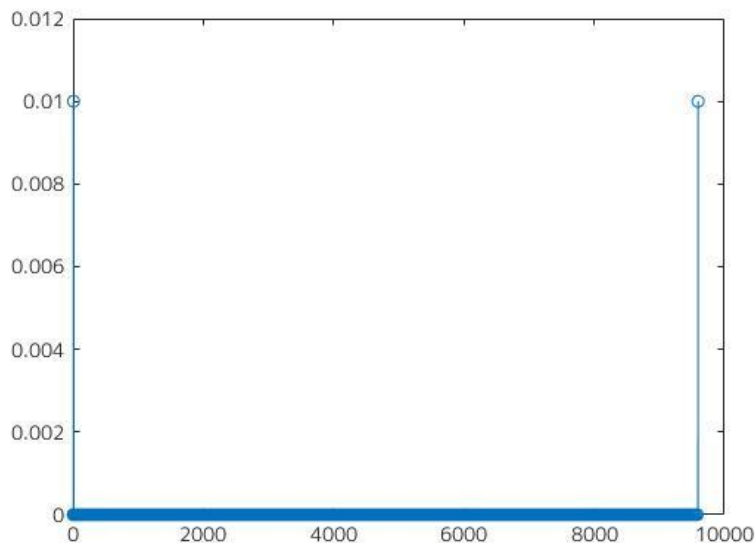
純音の時間波形



## 2-2. パワースペクトルの図示

下の左側の画像は0.6秒分のサンプル数9600個分すべてのパワースペクトルを図示したものである。始まりと終わりに強い値を持つが、詳細がわからない。

そこで、先頭の10個分を見てみると7点目におおきな値を得られていることがわかった。



# 問題2のまとめ

---

## 問題

純音(正弦波)のパワースペクトルを観察し, パワースペクトルから得られる情報と聴感的な印象の関係について考察せよ.

## まとめ

- ソースコードを示しながら, 純音の生成から, DFT(FFT)を用いたパワースペクトルの表示までの手順を説明した. 関数一つでパワースペクトルは表示できた
- パワースペクトルの強い点を示しているところを拡大して図示した

TODO: 2-3以降の小問題

# 問題4

---

じゃんけんの3手（「グー」「チョキ」「パー」）の単語発声を音声ファイルとして録音し、今後の音声認識の実験で利用できるよう準備せよ.

- いずれの音声ファイルも、サンプリング周波数 16 kHz, 信号長 600 ms とすること.
- 各パターンにつき、少なくとも2つは収録すること.



# 問題4の小問題

---

1. 音声収録の際に、注意したことや心がけたことがあれば、説明しなさい.
2. 記録した6つの音声（グー, チョキ, パーそれぞれ2つ）の時間波形を図示しなさい.
3. 記録した6つの音声（グー, チョキ, パーそれぞれ2つ）の対数パワースペクトルを図示しなさい. (TODO: 11/12以降に実施)
4. 時間波形や対数パワースペクトル等の「観察結果」や「考察」を示しなさい. (TODO: 11/12以降に実施)

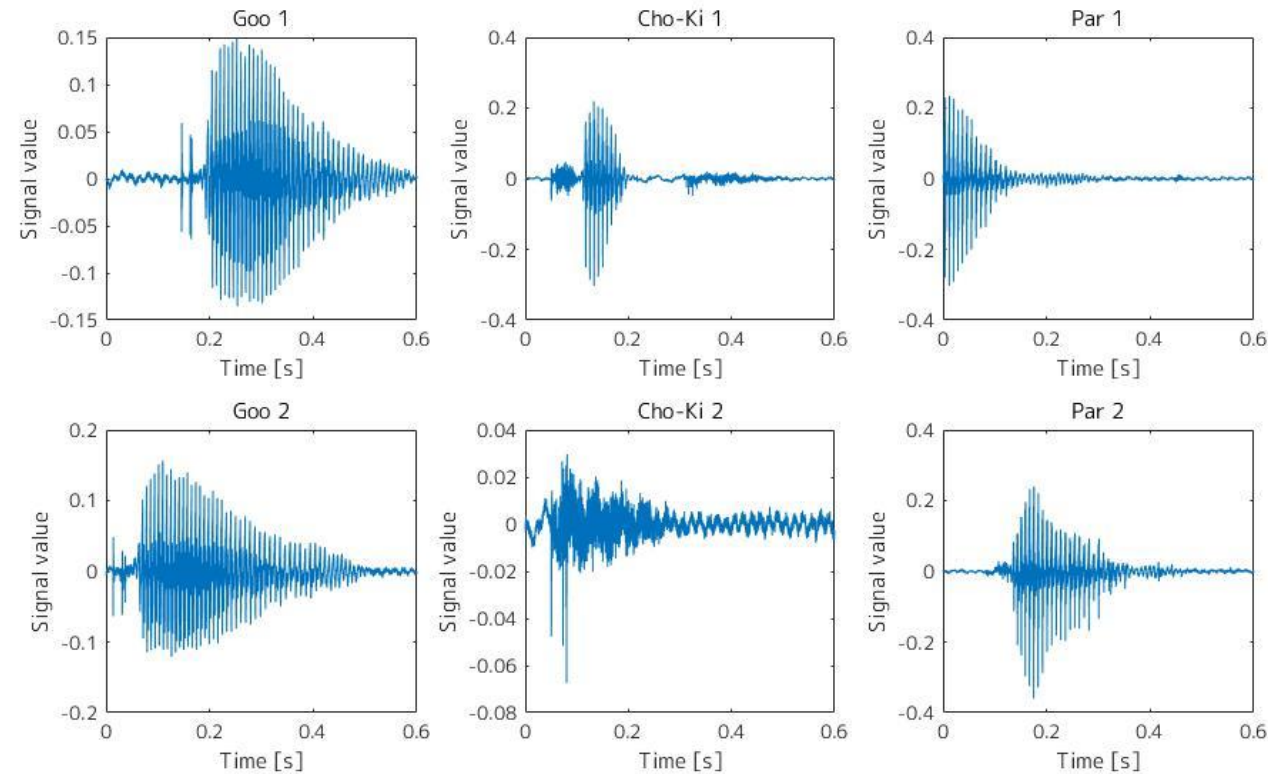
## 4-1. 音声収録の際に，注意したことや心がけたことがあれば，説明しなさい.

---

1. 該当コードに到達するとすぐに録音が始まるので，はじまるまえにpause(1)を挟んで録音待機をさせた
2. 0.6秒に収まるように発話タイミングを調整した
3. 音の大きさが振りきれて俗に言う音割れを起こさないように録音する音量を調整した

## 4-2.6つの音声の時間波形

MATLABのsubplot関数を使い、一度に6つの波形を表示させた



# 問題4のまとめ

---

## 問題

じゃんけんの3手(「グー」「チョキ」「パー」)の単語発声を音声ファイルとして録音し、今後の音声認識の実験で利用できるよう準備せよ.

## まとめ

MATLABを使いじゃんけんの3手それぞれ2つずつのサンプルを録音し、保存した  
それぞれの時間波形を示した

# 感想

---

MATLABでの音声録音がとても簡単に行えて驚きました。今回はUSBのインターフェースが一つしかなかったから大丈夫だったけど、複数ある場合にそれらを切り替えて録音することは可能なのかなと思った。

録音が目的だったら、家のほうがよかった。防音設備もあったし、マイク品質も貸し出されたものよりも良いので、ノイズの少ない良いサンプルが取れたはず。また、最初の1時間ネットにつながらずとても時間を無駄にした気分でした。

ただ、やはりオフラインでの安心感はあるなと感じました。