

# 音声処理実験 作業日報 第4回 SP-2

---

学生番号: 09430509

氏名: 今田将也

提出日: 2020年11月12日

# SP-2の概要

---

## 1. 問題2

1. 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た
2. 周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た.
  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. 問題3 (優先発展)

1. 適当な信号のスペクトログラムを表示して, その算出過程や結果について考察した

## 3. 問題4

1. 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？

青字は前回実施

## 2-3.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

### 1. 周波数軸(X軸)の換算

1. 周波数番号 $k$ に対応した周波数を持つベクトル $f_k$ を用意する.
2. MATLABの配列は1から始まるため,  $k=1+k'$ として用意する.

### 2. 片側化

1. サンプリング定理に従えば, サンプリング周波数 $A$  Hzにおいては,  $2/A$  Hzより大きい周波数成分は存在しない

### 3. 対数化

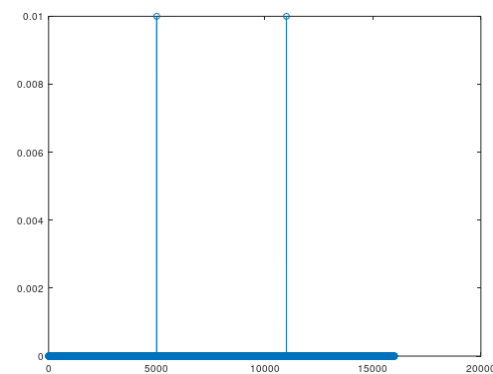
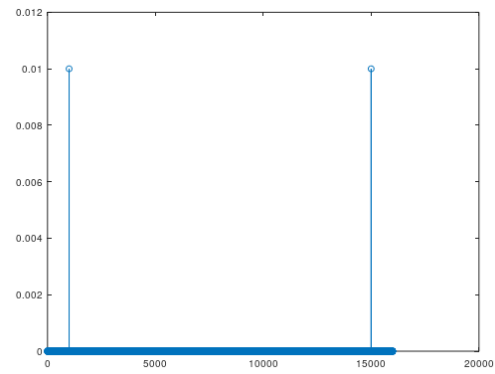
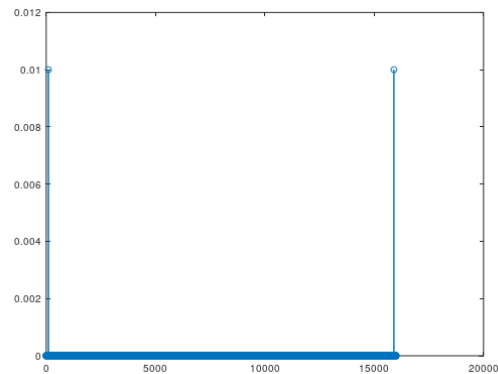
1. パワースペクトルは縦軸の数値を対数スケールに換算してdB(デシベル)という単位で示す
2.  $XPow\_dB = 10 * \log_{10}(XPow)$ という変換式を用いる

```
f_k = linspace(0, sampling_rate, signal_length_pt+1); % 16000を9600+1分割した配列
f_k(end) = []; % 最後消す
XPow((2+signal_length_pt/2):end) = []; % 折り返した後半を消して片側のみ
f_k = linspace(0, sampling_rate/2, 1+signal_length_pt/2); % 等分割
plot(f_k, XPow);
xlabel('Frequency [Hz]');
xlim([1 20]) % plot array index from 1 to 10
```

## 2-4. 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

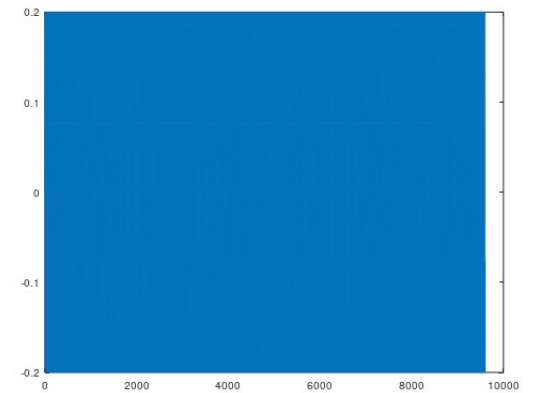
下図に示したのは左から100hz,1000hz,5000hzの正弦波のパワースペクトルの図である。

- サンプリングレートは16000hz
- 周波数が高くなるにつれ、パワースペクトルの特徴点の位置はナイキスト周波数である8000hzに集まっていくように見える



## 2-4. 正弦波信号の周波数と聴感的な印象

1. 実験に用いた16000hzのサンプリングレートのナイキスト周波数である8000hzでは音が聞こえなくなった
2. 周波数が高いと耳を劈くようなきつい音になる
3. 周波数が低いとボーッというようなこもった感じの音になる
4. 周波数が低いとはいえ、1秒間に200回もなっていると波形は見づらい
5. 周波数1000hzとかになると青いただの四角のグラフに見えた



# 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

- 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できた
- 周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、周波数の高い時、低いときについて考察した



# 問題3

---

適当な信号のスペクトログラムを表示して、その算出過程や結果について考察せよ.

# 問題3の小問題

---

1. 本問題の設定における番号(データ番号および周波数番号)と実尺度(時刻および周波数)の換算式を書きなさい.
  1. 換算表, あるいは, MATLABのコードとして書いても良い.
  2. 時刻を測る実尺度は s 単位(秒単位), または, ms 単位(ミリ秒単位)とすること.
  3. 周波数を測る実尺度は Hz 単位, または, kHz 単位とすること. (※角周波数ではない)
2. 適当な正弦波信号を作成し, その信号のスペクトログラムを表示し, 考察しなさい.

## 3-1. 変換式

---

横軸が時刻 $t$ に応じて変わる信号 $f(t)$ を, DFTを用いてパワースペクトルにすると, 横軸は周波数番号というものになる

そのときの関係式を以下に示す

$\Delta n = 1[pt] \leftrightarrow \Delta t = \frac{1}{F_S}[sec]$  1個あたりのデータ番号はサンプリング周波数の逆数の大きさになっている

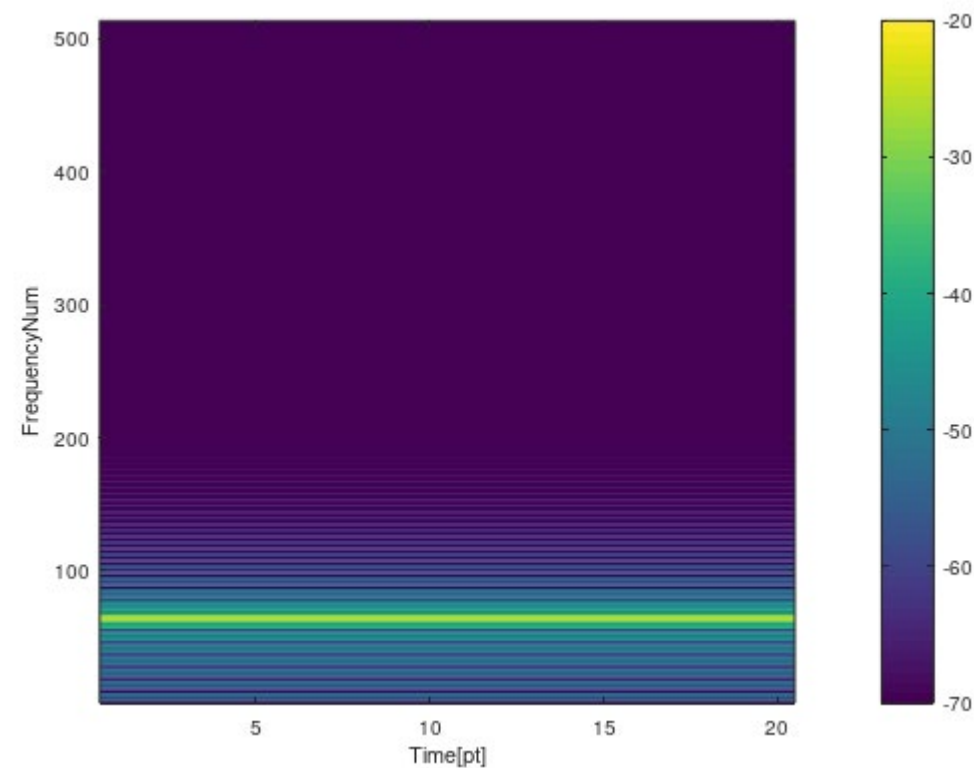
$\Delta k = 1[pt] \leftrightarrow \Delta f = \frac{F_S}{N}[Hz]$  周波数番号の1個あたりはサンプリング周波数をデータの長さの個数で割ったもの

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

```
X = fft(frame_x, fft_len) / fft_len; % 各フレームの信号に対するFFTの結果として, 1024x20の複素数の行列x  
Pow_X = abs(X) .^ 2; % 20個のFFTの結果を, 20個のパワースペクトルに変換  
Pow_X((2+fft_len/2):end, :) = []; % 後ろから(半分-1)個を消す=前から(半分+1)個を生かす 片側化  
Pow_X_dB = 10 * log10(Pow_X); % 対数化
```

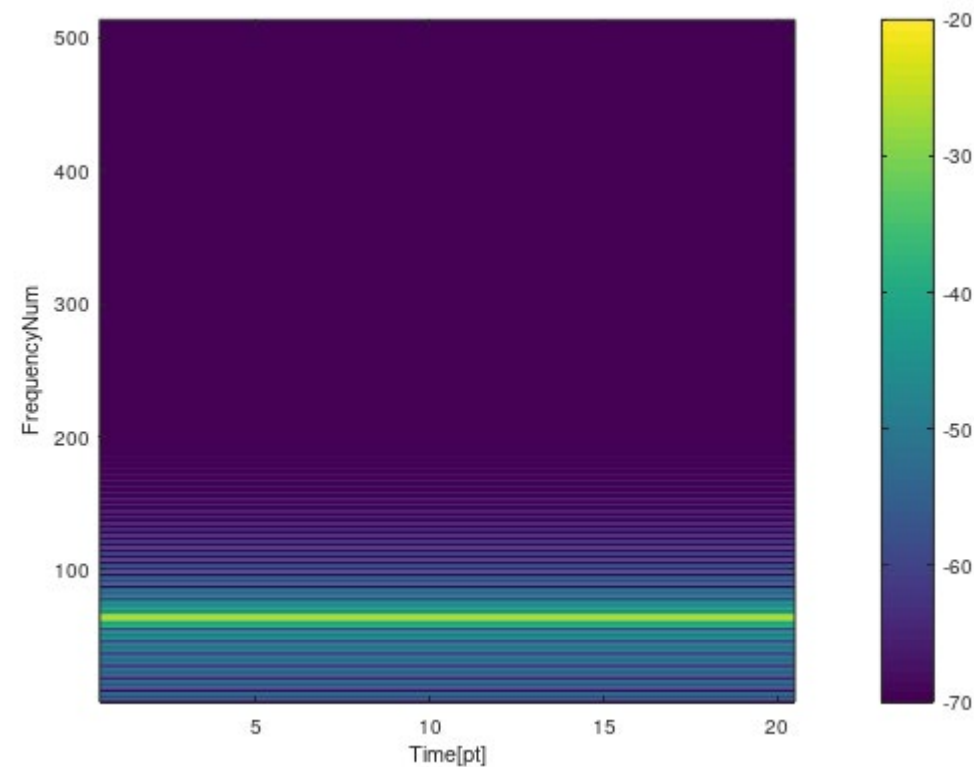
## 3-2.用いた尺度とスペクトログラム

- ✓ サンプリング周波数は 16 kHz (16,000 Hz)
- ✓ 信号長1秒
- ✓ 周波数1,000 Hzの純音
- ✓ 信号長が16000点
- ✓ 800点ずつに分割
- ✓ フレーム数は20個



## 3-2. スペクトログラムの考察

- ✓ 1000hzの正弦波を生成し、スペクトログラムを表示した.
- ✓ 純音であるため、ある一定の値のPowerが強く反応していることが見て取れる.
- ✓ おそらく、1000hzの周波数番号が黄緑に近い色の部分に格納されている
- ✓ 時間経過しても周波数のデータは損なわれない



# 問題3のまとめ

---

## 問題

適当な信号のスペクトログラムを表示して，その算出過程や結果について考察せよ．

## まとめ

正弦波の実データ(時間，周波数)を講義の設定された番号への変換を行うための式を示し，変換の算出過程のコードも示した

スペクトログラムを示し，時間と周波数の強さの特徴について考察した

# 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

---

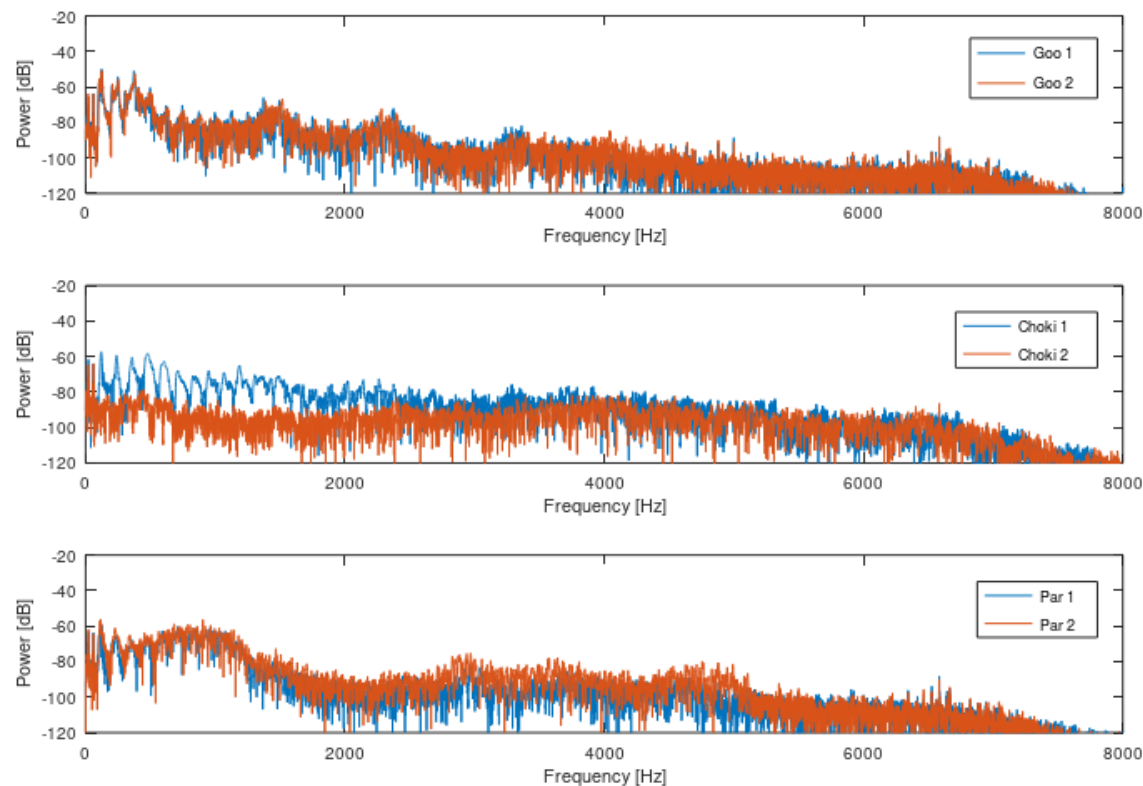
1. 音声収録の際に、注意したことや心がけたことがあれば、説明しなさい.
2. 記録した6つの音声（グー、チョキ、パーそれぞれ2つ）の時間波形を図示しなさい.
3. 記録した6つの音声（グー、チョキ、パーそれぞれ2つ）の対数パワースペクトルを図示しなさい.
4. 時間波形や対数パワースペクトル等の「観察結果」や「考察」を示しなさい.



## 4-3.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

右図に示す.

- ✓ 一番上がグーという発音2つの音の対数パワースペクトル
- ✓ 真ん中がチョキという発音2つの音の対数パワースペクトル
- ✓ 一番下がパーという発音2つの音の対数パワースペクトル



## 4-4. 「観察結果」や「考察」

---

### 観察結果

- グーとパーを見てみると、殆どの周波数の部分で2つの音声のパワースペクトルに一致している
- チョキは4000Hz以下の部分でズレが大きい
- 2000hz以下の部分でのパワーが大きい

### 考察

- 発話者の声が低いため、低周波数域のパワーが強くなっていると思う
- チョキがずれている理由は恐らく発音の問題だろう. もしくは同じ周波数域でのノイズが混在している可能性がある

# 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

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しめし、それぞれの観察結果および考察について述べた

TODO::小問題(優先発展♪)

記録した6つの音声(グー, チョキ, パーそれぞれ2つ)のスペクトログラムを図示しなさい.

# 感想

---

問題3のスペクトログラムについて, どのタイミングでどのような単位の軸になっているかがわからなかった.

また, 問題4では現実世界の対数パワースペクトルを表示したところ正弦波よりも面白い特徴を見ることができた. スペクトログラムの表示は次回行いたい.

次回すること

問題4のスペクトログラム表示

問題5以降