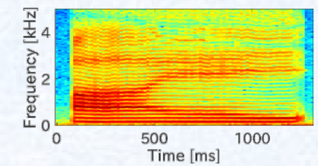


音響基礎の基礎

鳥谷 輝樹

山梨大学 大学院総合研究部 工学域

Email: t.toya@yamanashi.ac.jp

目次

音と情報技術

- 音は情報を伝える媒体（メディア）
- **コンピュータの登場と進化**

↓

音情報を精密かつ効率的に抽出・表現できるようになった！

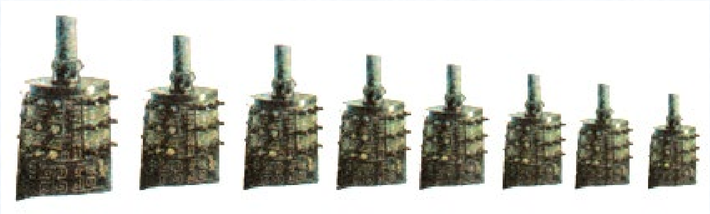
- **音響学**は**物理・数学**を基礎にして，今や高度な**計算機技術**を駆使する学問に

↓

- 何から音が鳴ったのか？（イベント検知）
- 音をどのように伝えるか（音響設計）
- どんな感情を含むか？（音声認識）
- ヒトにはどう聞こえているか？（心理）
- 邪魔な音をどうやって消すか？（雑音除去）

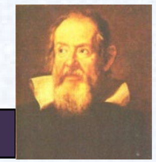
音響学の歴史（はじまり）

- 古代の人も**音を情報媒体として利用していた**
  - 音声, 音楽, 警報などの伝送手段
  - しかし,「音は空気中の波である」とは知らなかった
  - 音響学は 科学（Science）ではなく 哲学（Philosophy）だった



音響学の歴史（哲学 → 物理学）

■ 17世紀の終わりに 物理音響学 が確立



1564 G.Galilei 1642

1588 M.Mersenne 1648

1608 E.Torricelli 1647

1623 B.Pascal 1662

1627 R. Boyle 1691

1629 C. Huygens 1695

1635 R.Hooke 1703

1642 I. Newton 1727

音は空中の波動現象！

音響学の歴史（物理学→工学→情報学）

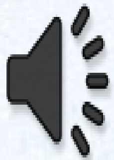
■ 物理音響学の熟成

- ➢ **J. Fourier（1768－1830）**：フーリエ級数展開・フーリエ変換
- ➢ H. v. Helmholtz（1821－1894）：聴覚の研究

■ 音響工学の発展

- ➢ A. G. Bell（1847 ~ 1922）：電話の発明
- ➢ T. Edison（1847 ~ 1931）：蓄音機の発明

■ 情報学との融合

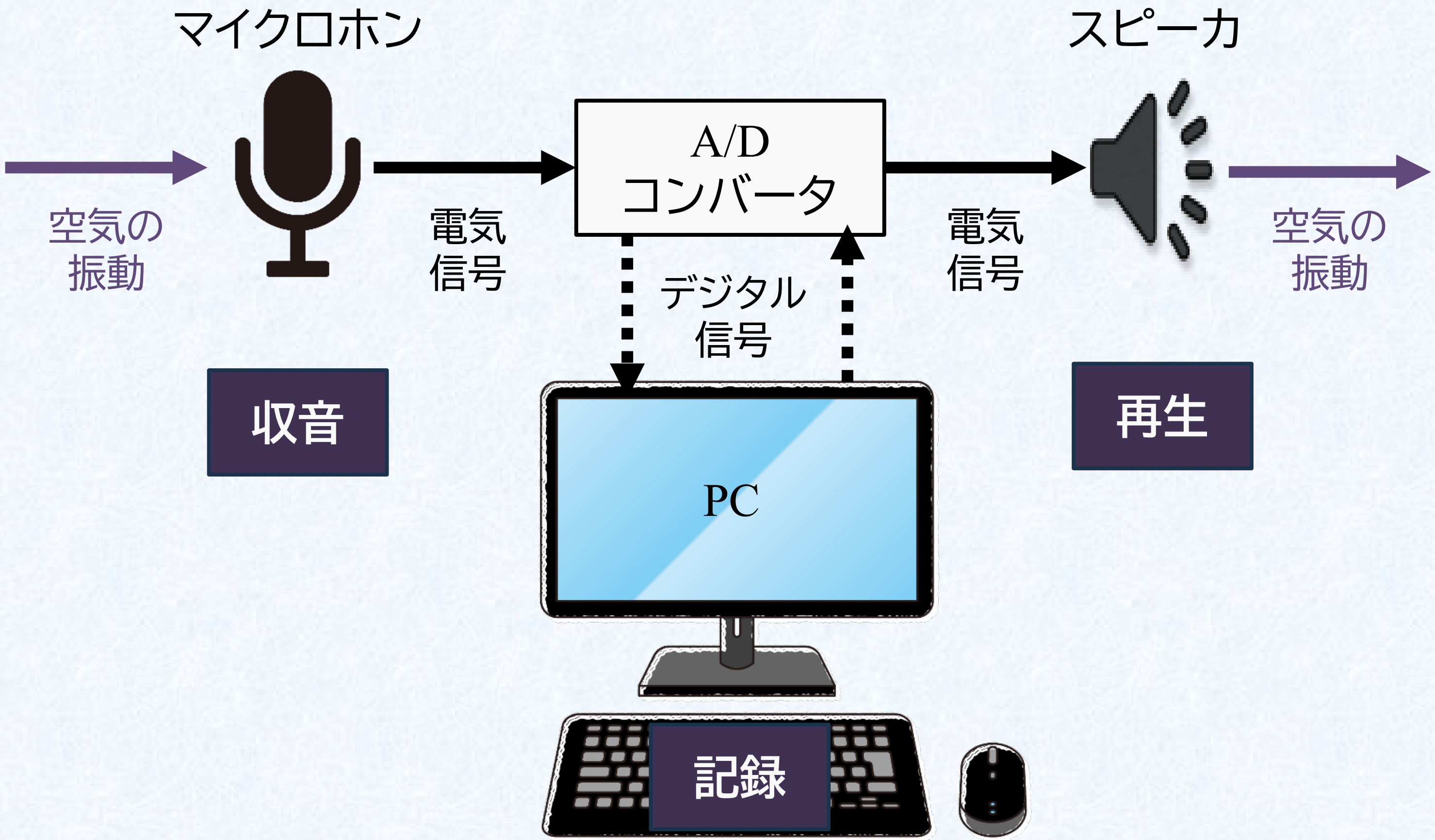


- ➢ C. E. Shannon (1916 ~ 2001)：情報理論
- ➢ J.W.Cooley & J.W.Tukey（1965）：高速フーリエ変換アルゴリズム

音がコンピュータで手軽に扱えるようになった！

1. 音の物理と数学

収音・記録・再生



物理としての「音」

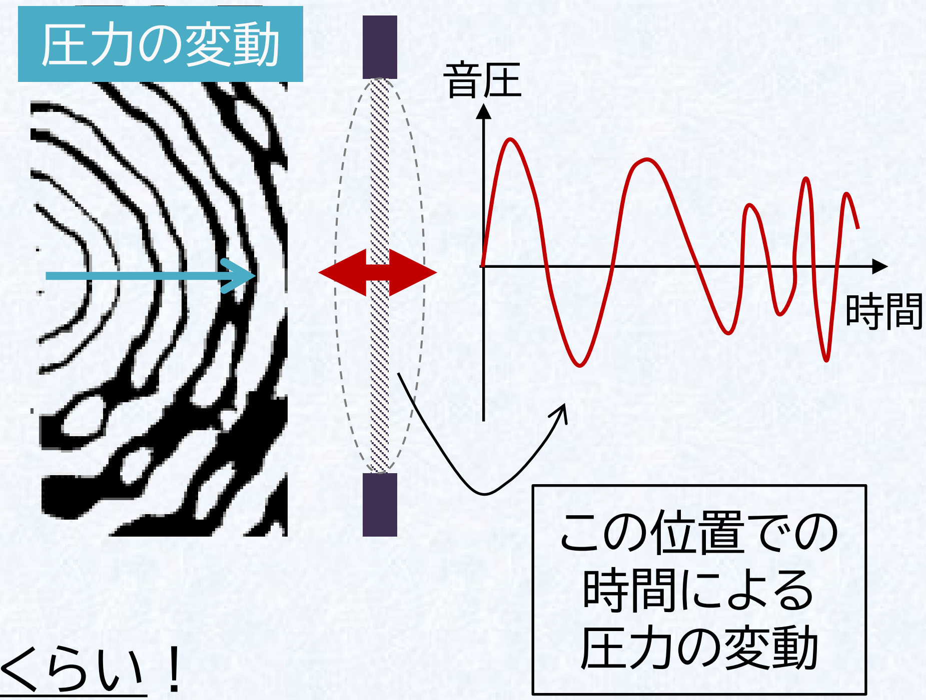
■ 空気の微小な圧力変動

➢ 「**音圧**」という
（圧力なので 単位は Pa）

➢ 大気圧（約 10^5 Pa）に対する
圧力の増減の繰り返し
（増減しないと「音」にならない！）

➢ ふだん聞く音の音圧は
- 会話で 10^{-2} Pa 程度
- イヤホン聴取で 10^{-1} Pa 程度

→ 大気圧の1,000,000 分の 1 くらい！



音圧は桁扱いが大変 → レベル表現

- 音圧レベル $L = 10\log_{10}\frac{p^2}{p_0^2} = 20\log_{10}\boxed{\frac{p}{p_0}}$ $(p_0 = 20\ \mu\mathrm{Pa})$

比

音圧 [Pa]	音圧レベル [dB]
20	120
1	94
0.2	80
0.02	60
2×10^{-3}	40
2×10^{-4}	20
2×10^{-5}	0

ライヴ会場

非常警報音 (1 m)

会話の音声

静かなレストラン
図書館
静かな音楽ホール

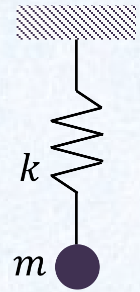
最小可聴値 (2 kHz)

最も単純な音の表し方：正弦波

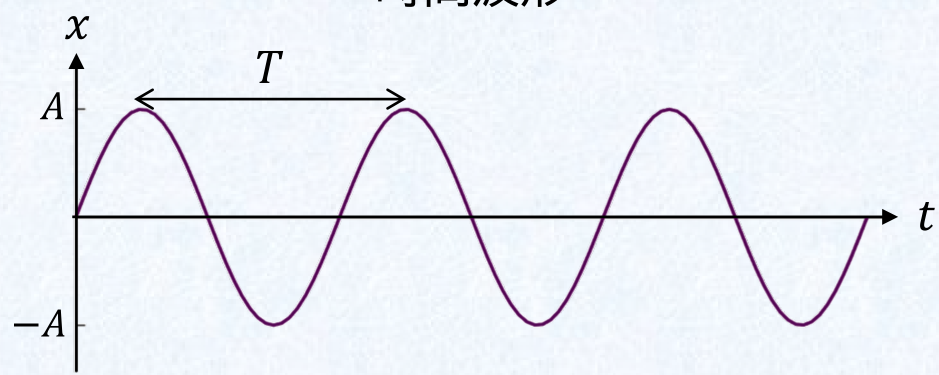
- ばねとおもりの振動
  - A : 振幅
  - T : 周期 [s]

$$x(t) = A\sin\frac{2\pi}{T}t$$

（時間の関数）



時間波形



- 音波も考え方は同じ
  - $x(t)$ を 音圧 と考えればよい（マイク・スピーカでは「電圧」になるが）
  - 音では $f = 1/T$ を 周波数（単位は Hz）と呼び
    $x(t) = A\sin 2\pi f t$ とすることが多い（高校物理では「振動数」）

コンピュータでの音信号の表現

- コンピュータではディジタル信号(≒ 数列)として扱う
  - **標本化** (サンプリング):

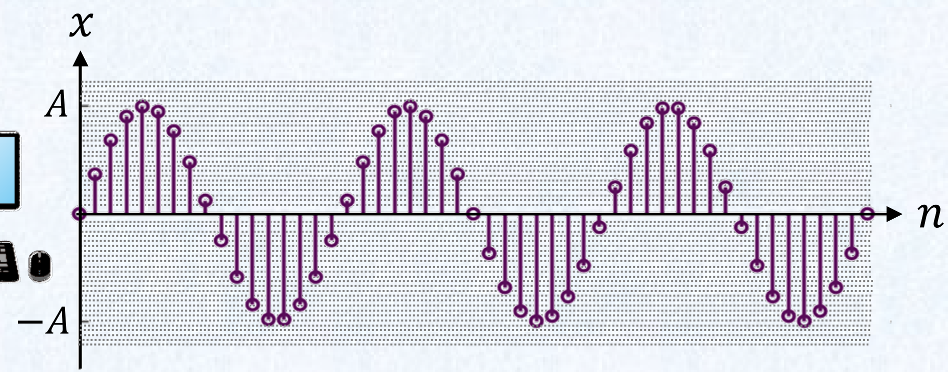
    とびとびの時間間隔 T_s ごとに信号の値を取得して数列にする

    $$x_n = A \sin 2\pi f T_s n$$

  - **量子化**:

    ある決まった精度で数列の各項の値を整数近似する

    → 最終的には2進数のビット列



n : $[0, 1, 2, \cdots, N]$ (インデックス)

$T_s n$: $[0, T_s, 2T_s, \cdots, NT_s]$

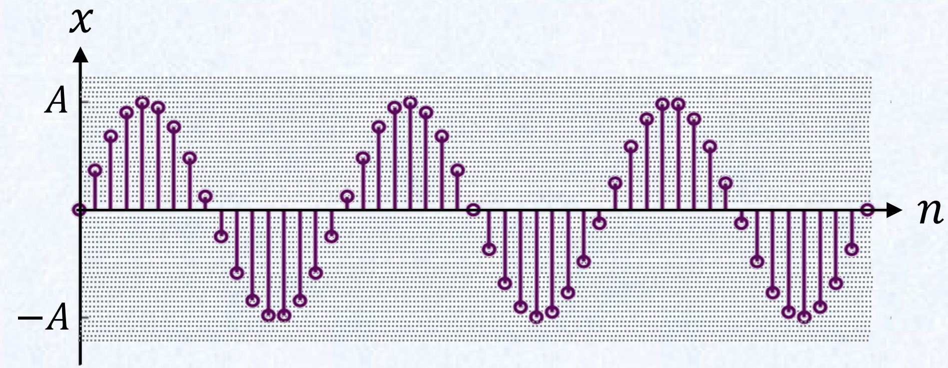
(初項 0, 公差 T_s の等差数列)

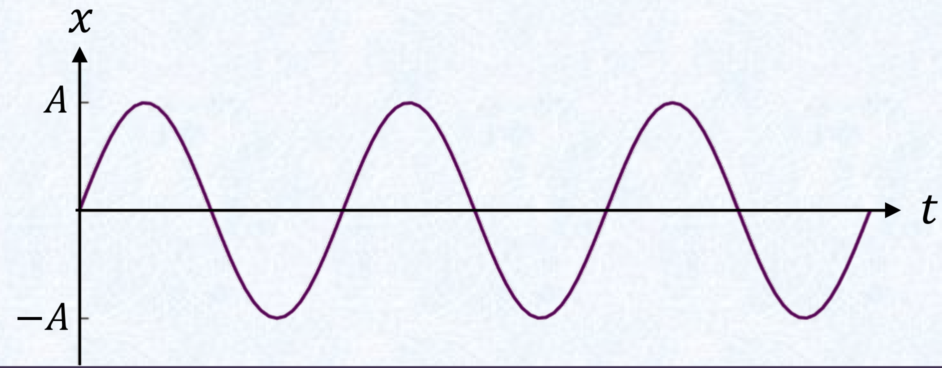
x_n : $[x_0, x_1, x_2, \cdots, x_N]$

0	1	1	…	0	
1	0	1	…	0	2 進数
0	1	0	…	1	

コンピュータでの音信号の表現(続)

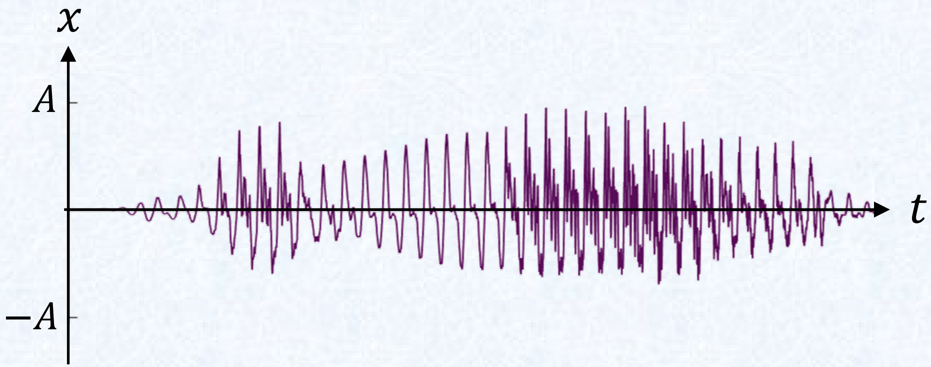
- 以後の説明・演習では「連続的に見える波形」を図示
  - 実態はこんなデータの列ですが…
  - 連続量のほうが見やすいので, 説明ではこう表します





実際の音はもっと複雑

- 振幅・周期（周波数）はつねに変化している
  - あらゆる音に対して
    音圧の時間変化を
    シンプルな関数で
    記述するのは困難……



- 数学を駆使して解析してみる

2. フーリエ変換・スペクトル

フーリエ級数

- J. Fourier（1768－1830）

「任意の周期関数は三角関数の級数で表すことができる」

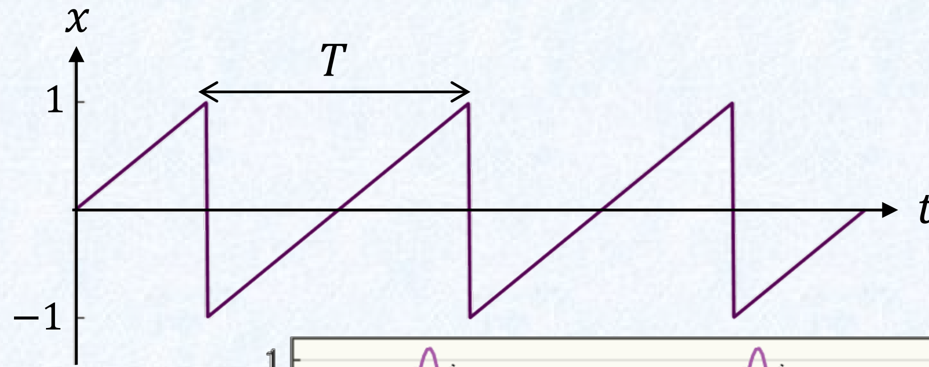
$$x(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos n\omega t + b_n \sin n\omega t)$$

$$\left(a_n = \frac{2}{T} \int_0^T x(t) \cos n\omega t \, dt \, , \qquad b_n = \frac{2}{T} \int_0^T x(t) \sin n\omega t \, dt \, , \qquad \omega = \frac{2\pi}{T} \right)$$

➢ 性質の不明な関数であっても, 異なる周期の三角関数（cos と sin）に分解して性質を知ることができる！

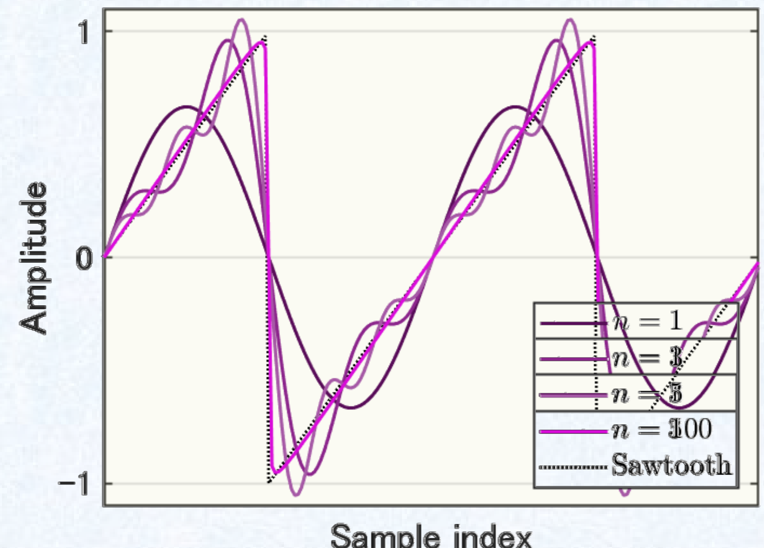
例①:のこぎり波のフーリエ級数展開

- のこぎり波:
  - $x(t) = \omega t \quad \left(-\frac{T}{2} \leq t < \frac{T}{2}\right)$
    $x(t) = x(t + nT)$



- フーリエ級数展開:

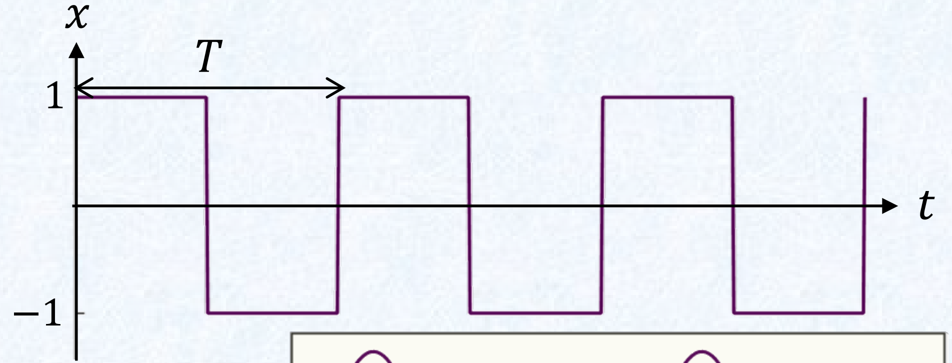
$$x(t) = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{2}{n} \sin n\omega t$$
$$= 2\sin\omega t - \sin 2\omega t + \frac{2}{3}\sin 3\omega t + \cdots$$



例②：矩形波のフーリエ級数展開

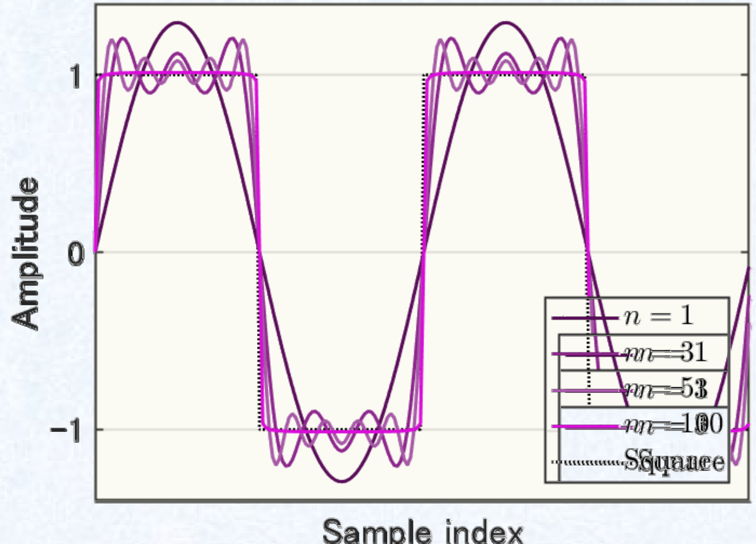
■ 矩形波：　$x(t) = x(t + nT)$

➢ $x(t) = \begin{cases} 1 & \left(-\frac{T}{2} < t < 0\right) \\ -1 & \left(0 < t < \frac{T}{2}\right) \\ 0 & \left(t = -\frac{T}{2},\ 0,\ \frac{T}{2}\right) \end{cases}$



■ フーリエ級数展開：

$$x(t) = \sum_{n=1}^{\infty} \frac{4}{(2n-1)\pi} \sin(2n-1)\omega t$$
$$= \frac{4}{\pi}\sin\omega t + \frac{4}{3\pi}\sin 3\omega t + \frac{4}{5\pi}\sin 5\omega t + \cdots$$



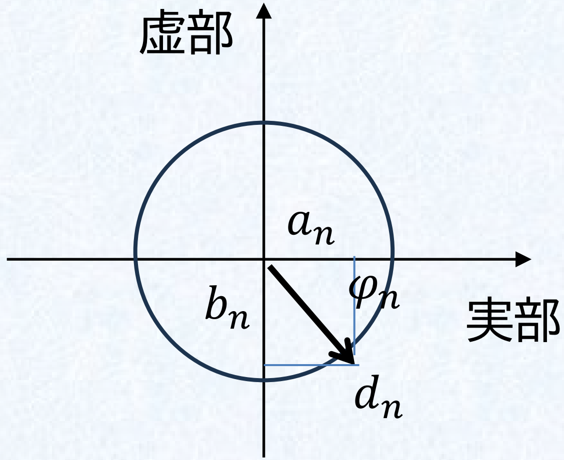
もっと数学的に扱いやすく…！

$$x(t) = \frac{a_0}{2} + \sum_n (a_n \cos n\omega t + b_n \sin n\omega t)$$

- cos と sin を考えるのは面倒くさい…
  - $a_n \cos n\omega t + b_n \sin n\omega t$

  $= \sqrt{a_n^2 + b_n^2} \cos(n\omega t - \varphi_n)$ （φ_n: 偏角）

  （どこかで見たことが…？）



- 複素数 $d_n = a_n - jb_n$ （$j = \sqrt{-1}$）を考えてみると
  - 絶対値 $|d_n| = \sqrt{a_n^2 + b_n^2}$
  - 偏角 $\arg d_n = \varphi_n$

$\sqrt{a_n^2 + b_n^2}$ (とφ_n)の値さえ分かればよい.
それなら, 複素数 d_n みたいに扱うほうがスッキリ記述できる！

複素フーリエ級数展開

$$x(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos n\omega t + b_n \sin n\omega t)$$

$$= \underbrace{\boxed{\frac{a_0}{2}}}_{c_0} + \sum_{n=1}^{\infty} \left(\underbrace{\boxed{\frac{a_n - jb_n}{2}}}_{c_n} e^{jn\omega t} + \underbrace{\boxed{\frac{a_n + jb_n}{2}}}_{c_{-n}} e^{-jn\omega t} \right)$$

$$= \sum_{n=-\infty}^{\infty} c_n e^{jn\omega t}$$

複素正弦波 という

オイラーの定理：

$$e^{j\theta} = \cos\theta + j\sin\theta$$

を用いて式変形

係数 c_n は

$$\boxed{c_n = \frac{1}{2\pi} \int_{-\infty}^{\infty} x(t) e^{-jn\omega t} dt}$$

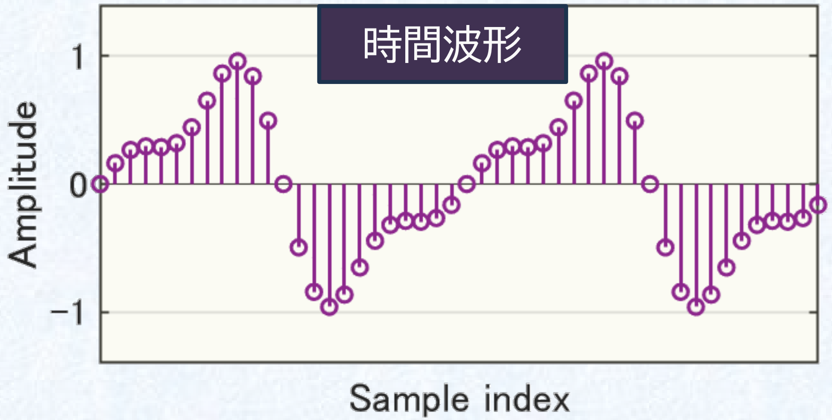
- 音信号 $x(t)$ を（複素）正弦波に分解できた！
  - 各 n について，正弦波の成分の振幅 $|c_n|$ を表した図を「（振幅）スペクトル」という

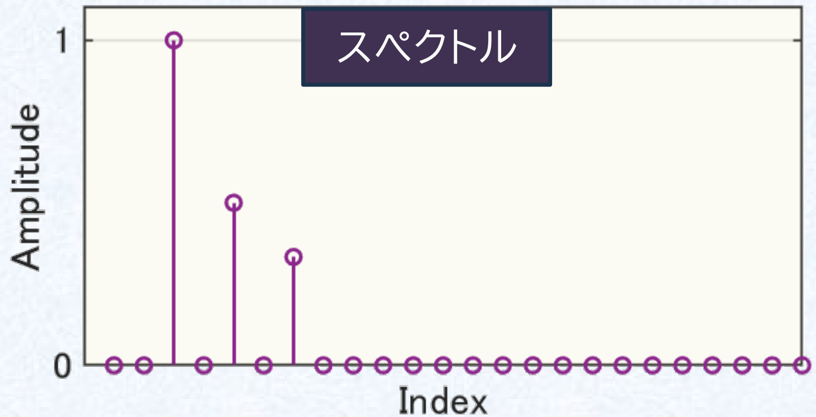
ディジタル信号のフーリエ変換

- 関数の積分 → 数列では「和」
  - $X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt$

$$X_k = \sum_{n=0}^{N-1} x_n e^{-j\frac{2\pi}{N}kn} \quad \left(\begin{array}{c} N\text{ 点の} \\ \text{有限な} \\ \text{数列} \end{array}\right)$$

  - ディジタル信号 x_n をフーリエ変換してスペクトル $|X_k|$ を得た（右図）

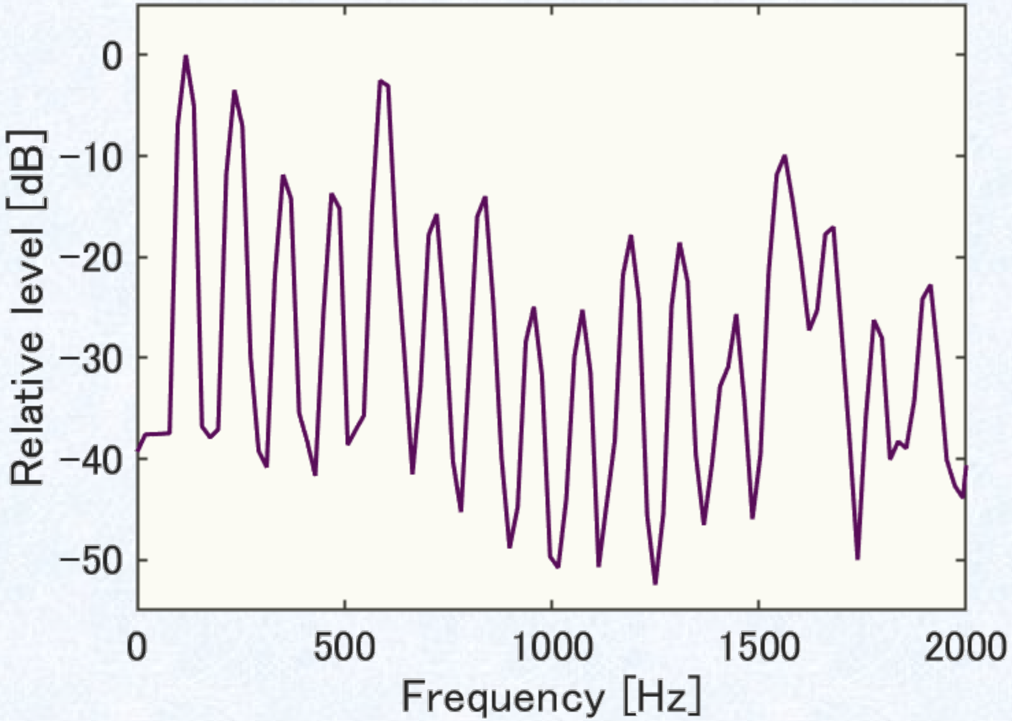




音のスペクトル

- 周波数 f ごとの振幅(＝成分の強さ)の分布を表す
  - 「音圧」で見ると振幅の変動範囲が広すぎる

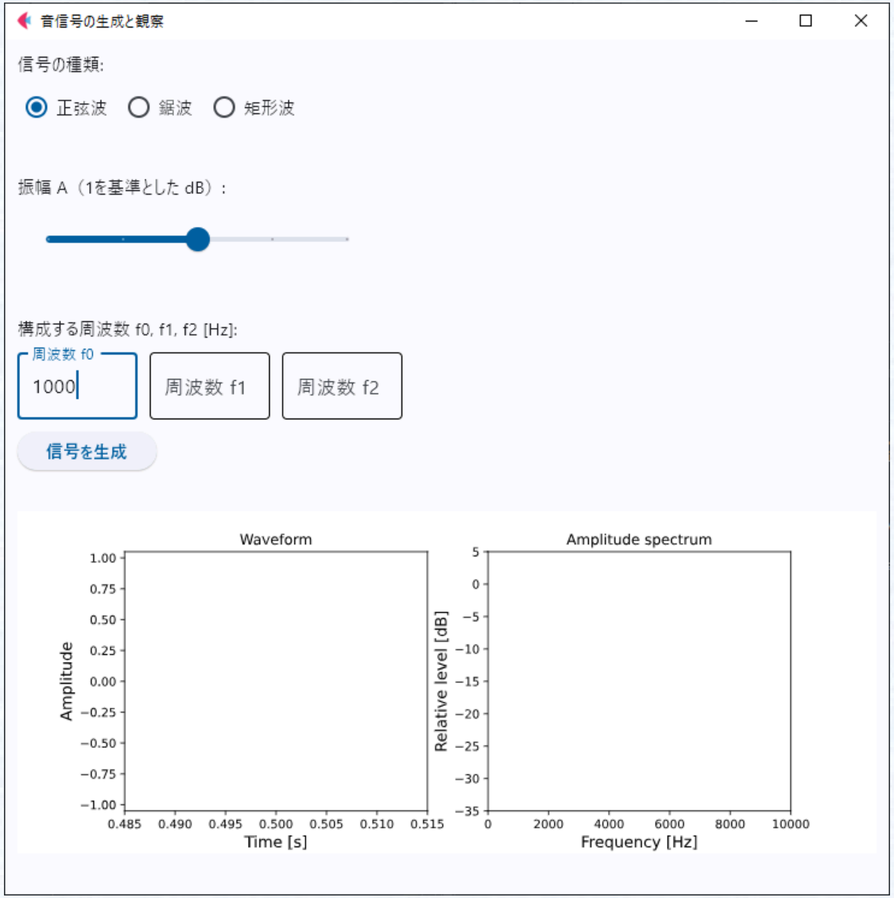
    ある基準を決めて，基準に対する比を「レベル」(dB)で表示することが多い



演習① 音を生成して観察しよう

- ターミナルで下記を実行
  - `python3 Ex1.py`
- GUI の操作
  - 信号の種類（正弦波・鋸波・矩形波）を選択する
  - 振幅のスライドバーを，まずは真ん中（−12 dB）くらいにしてみる
  - 周波数を入力するボックスに値を入力（200 Hz 以上, 5,000 Hz 以下くらいがよい）
  - 「信号を生成」ボタンを押して，
    - ヘッドホンから音が聞こえる
    - 時間波形とスペクトルが表示される

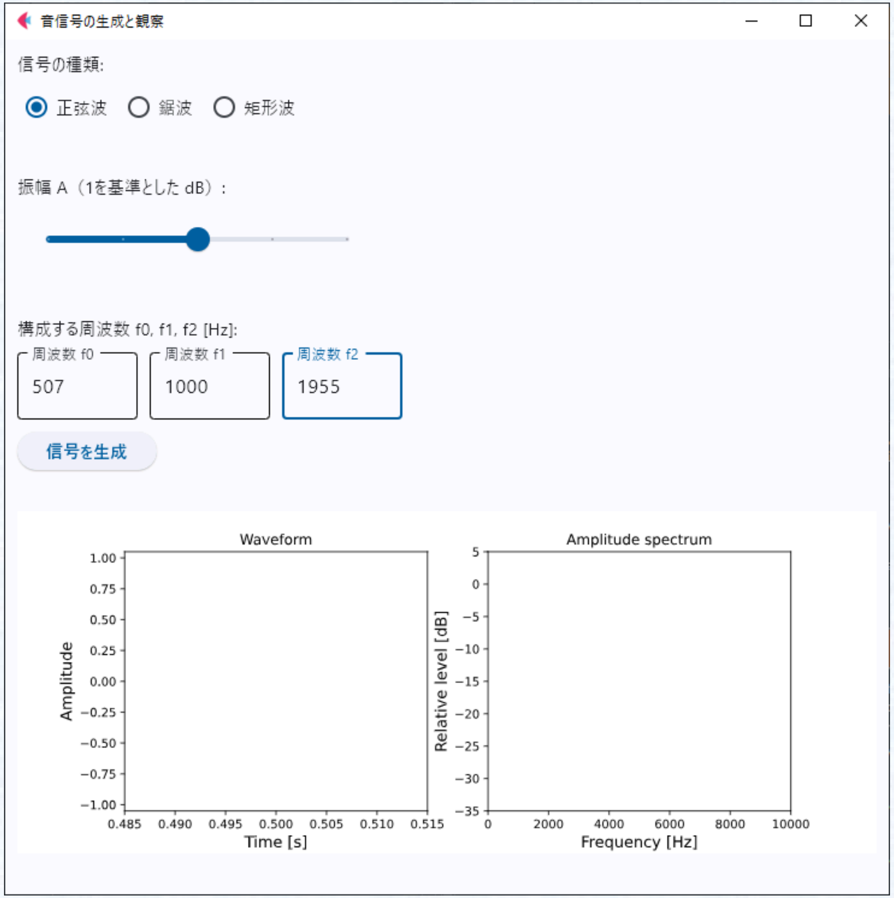
    ことを確認する.



演習① 音を生成して観察しよう(続)

- 調べてみよう
  - 正弦波で, 周波数 f0 だけを入れてその値を高くしていく(例. 500→800→1100)と, スペクトル・音の聴こえはどう変わるか？
  - 正弦波・鋸波・矩形波で, 周波数 f0 がまったく同じ(例. 1000)とき, スペクトルと音の聴こえはどうか？
  - 正弦波で, 周波数 f0, f1, f2 を
    - 1倍, 2倍, 3倍 の関係
    - 全部テキトーな数字

    とした場合はどうか？



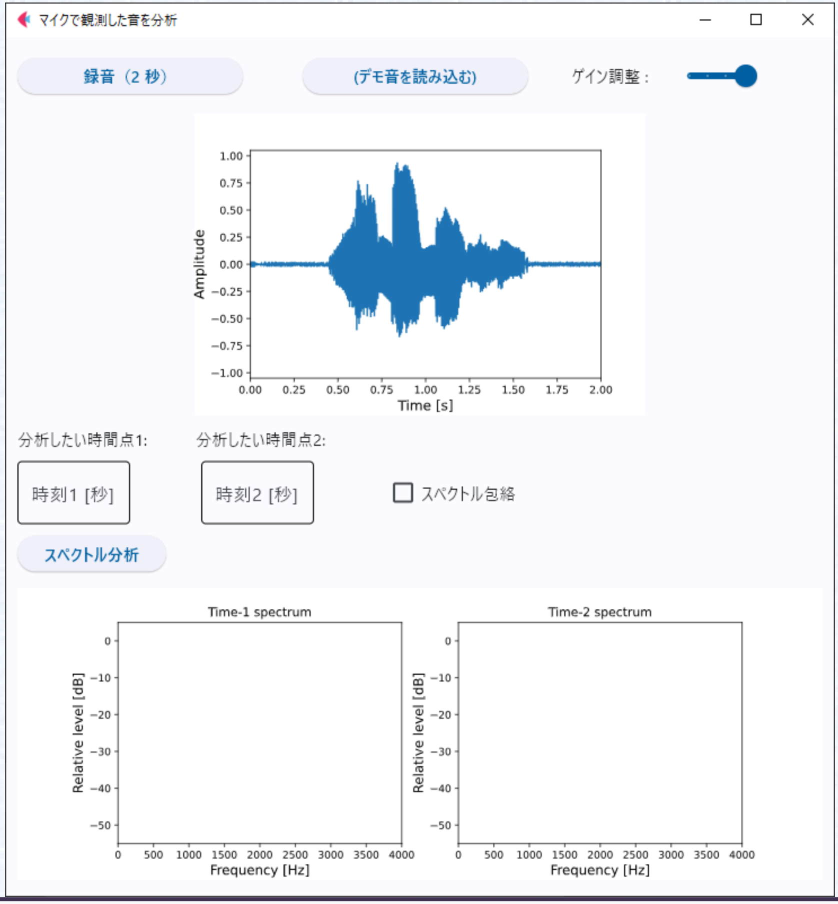
演習② 録音した声を観察しよう

■ ターミナルで下記を実行

- `python3 Ex2.py`

■ GUI の操作

- マイクスイッチをONにする
- 録音ボタンを押す
  （ピーと聞こえたら一息おいて，はっきりしゃべる）
  （2秒間しかないので，短い単語などがよい）
- 時間波形を見て，観察してみたい位置の時間を小数で入力してみる（2か所）
- 「スペクトル分析」ボタンを押すと2か所の時間位置でのスペクトルが出る
  「スペクトル包絡」チェックを入れてから，もう一度分析ボタンを押すと，スペクトルの「概形」が見える



波形・スペクトルと「音の感覚」の関係

■ 音の基本的属性（感覚の三要素）

➢ **音の大きさ**

小さい⇔大きい と感じる感覚量
（cf）スマホの音量調整ボタン

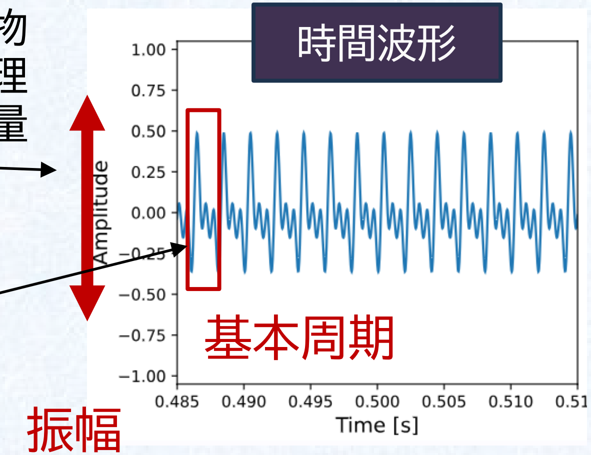
➢ **音の高さ**

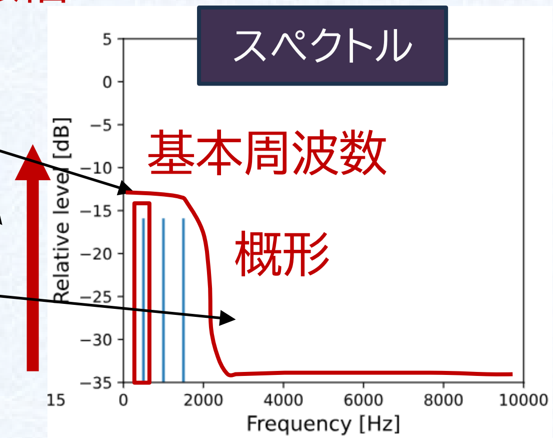
低い⇔高い と感じる感覚量
（cf）ピアノの鍵盤位置

➢ **音色(ねいろ)**

「音の大きさ」「音の高さ」以外の
音の感覚すべて

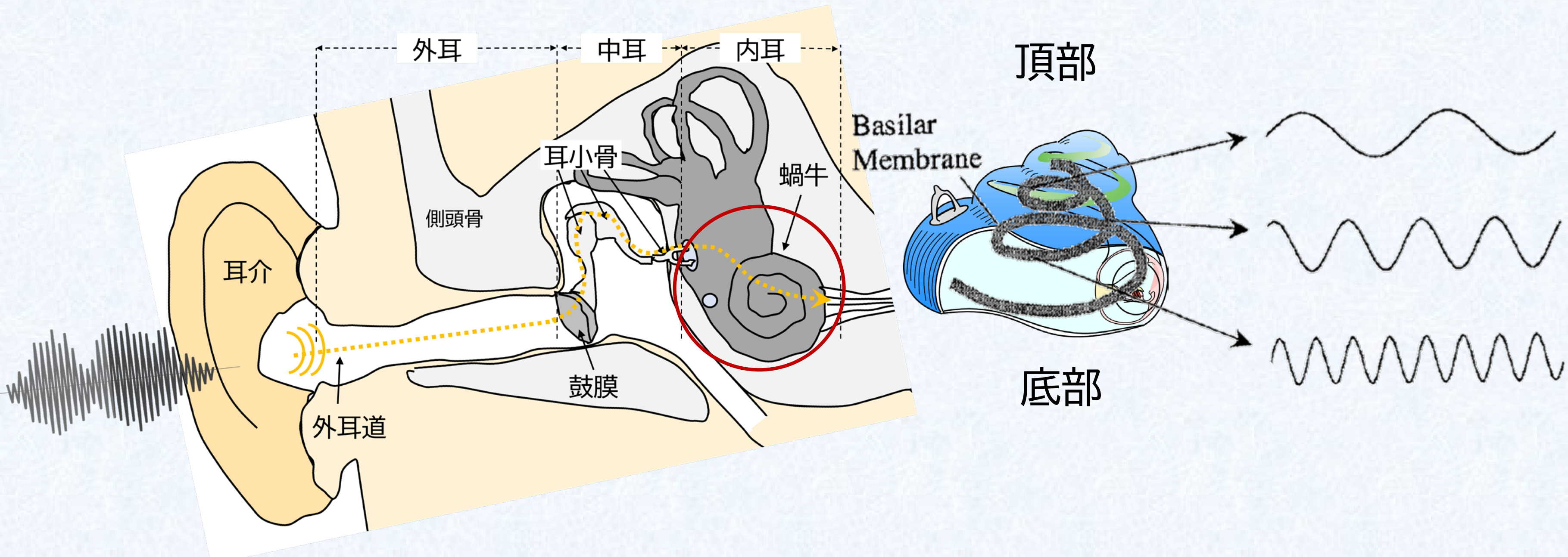
物理量





ヒトの耳も音を分析している

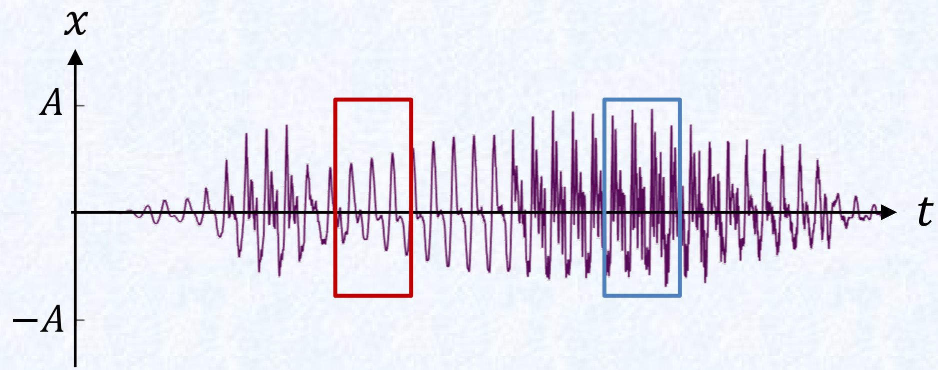
- 蝸牛(かぎゅう)は高度な周波数分析器



3. 変化する音の分析と応用

ヒトの声は短い時間で変化する

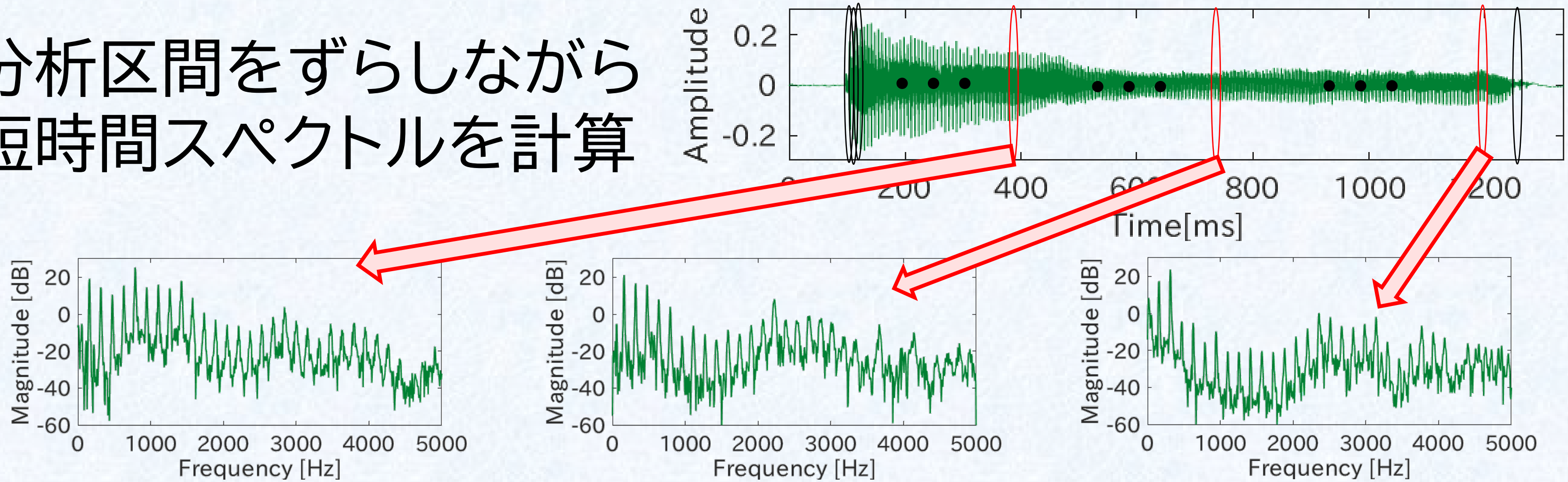
- 波形の切り取る位置で スペクトルの分析結果は 変わる……
（演習2で体験した）



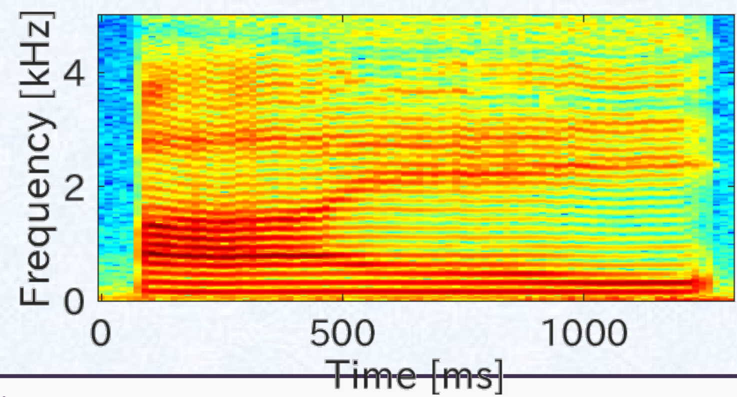
- 時間波形みたいに，
スペクトルが時間とともに変化している様子を
可視化できないだろうか……？

短時間分析 と スペクトログラム

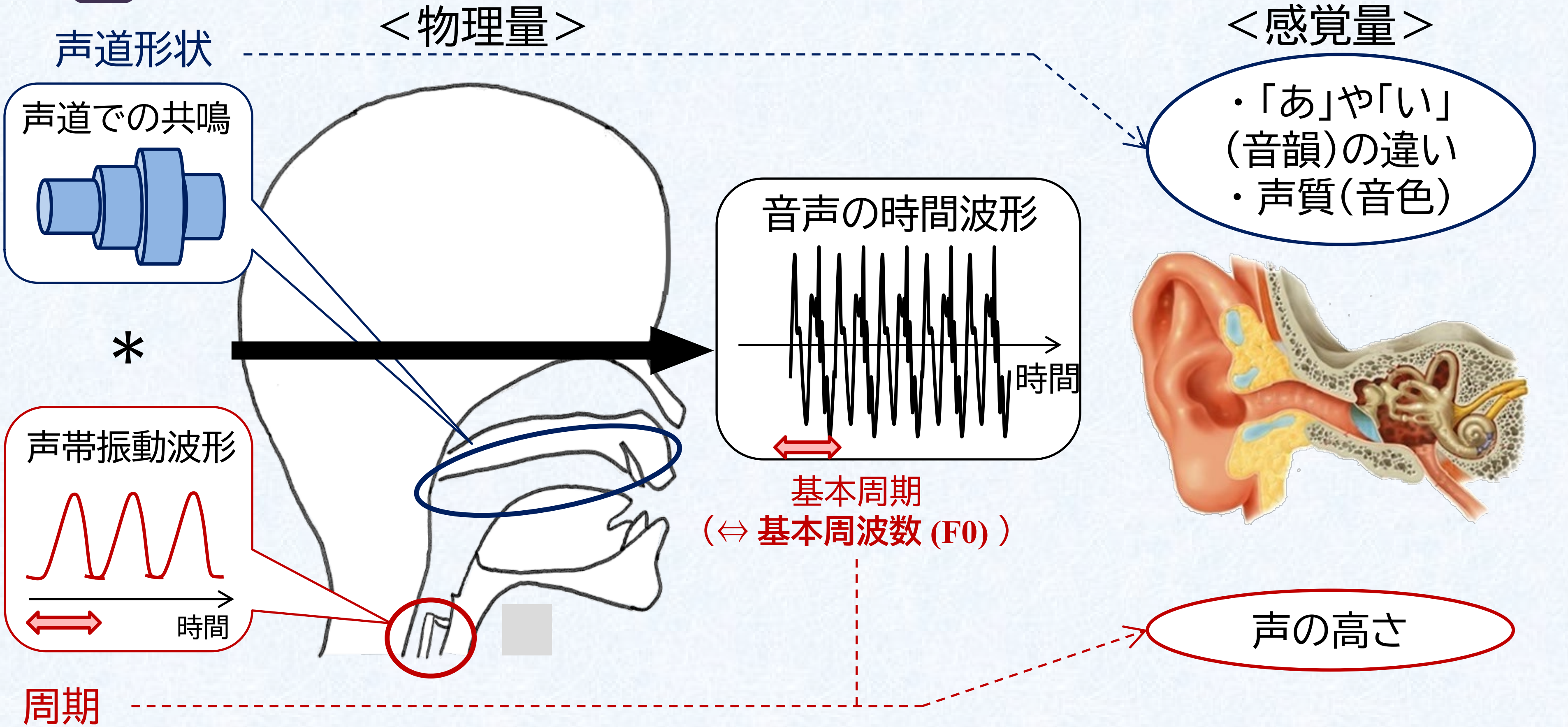
- 分析区間をずらしながら短時間スペクトルを計算



- 分析区間ごとのスペクトルを色の違い(濃淡)で表示したものが<u>スペクトログラム</u>

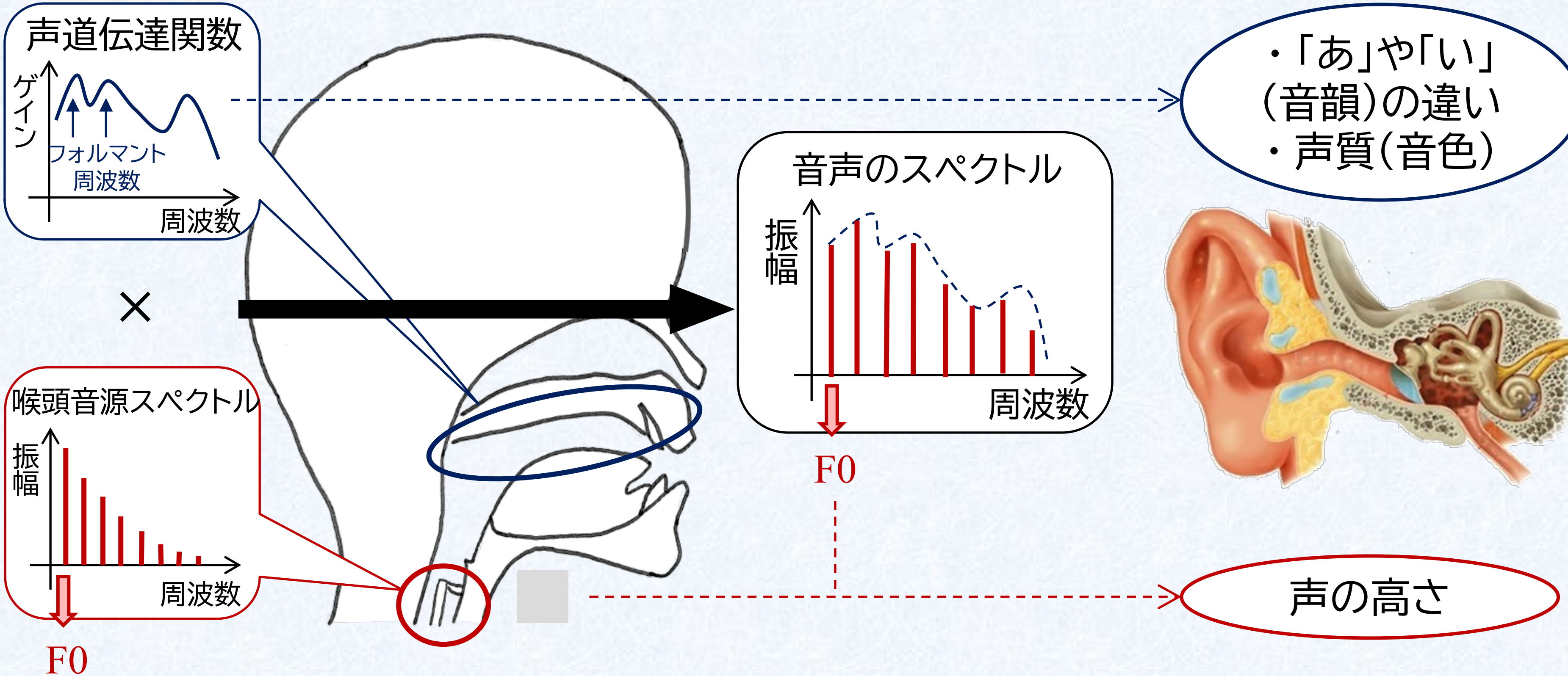


音声発話のメカニズム



＜物理量＞
＜感覚量＞
声道形状
声道での共鳴
・「あ」や「い」
（音韻）の違い
・声質（音色）
音声の時間波形
時間
*
声帯振動波形
時間
基本周期
（⇔ 基本周波数 (F0) ）
声の高さ
周期

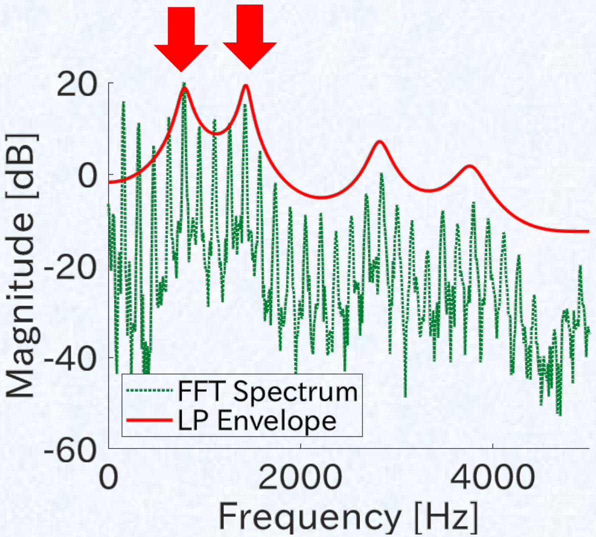
音声のスペクトル構造

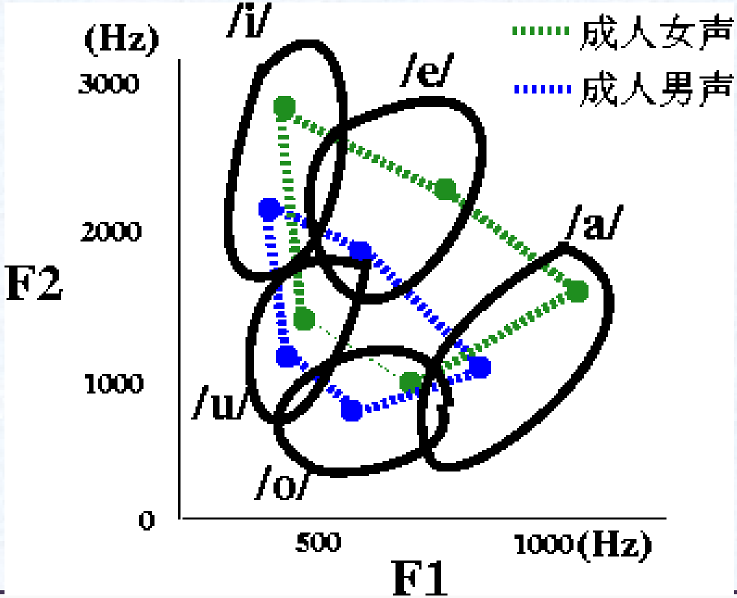


声道伝達関数
ゲイン
フォルマント
周波数
周波数
×
喉頭音源スペクトル
振幅
周波数
F0
音声のスペクトル
振幅
周波数
F0
・「あ」や「い」
（音韻）の違い
・声質（音色）
声の高さ

母音の共鳴特性の違い

- スペクトルのピークに違いが現れる
  - 最も低い周波数域にあるピーク(第一フォルマント)の周波数(F1)
  - 次に低い周波数域にあるピーク(第二フォルマント)の周波数(F2)

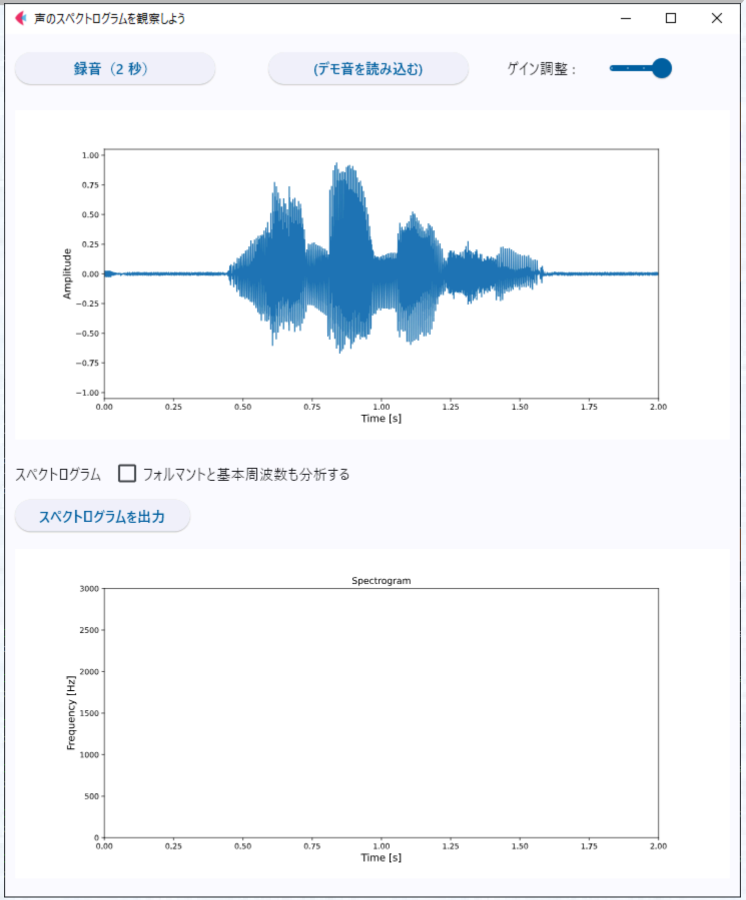




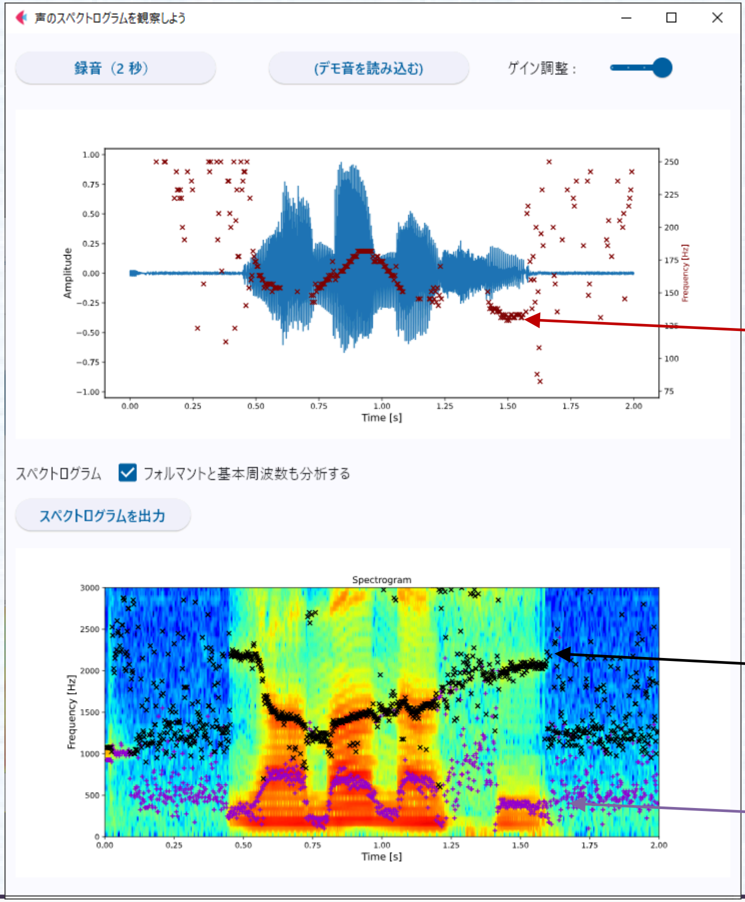
http://media.sys.wakayama-u.ac.jp/kawahara-lab/

演習③ 声の色々な特徴を観察しよう

- ターミナルで下記を実行
  - python3 Ex3.py
- GUI の操作
  - 録音ボタンを押す
    （ピーと聞こえたら一息おいて, はっきりしゃべる）
    （2秒間しかないので, 短い単語などがよい）
  - 時間波形を見て, 観察してみたい位置の時間を小数で入力してみる（2か所）
  - 「スペクトル分析」ボタンを押すと2か所の時間位置でのスペクトルが出る
    「スペクトル包絡」チェックを入れてから, もう一度分析ボタンを押すと, スペクトルの「概形」が見える



演習③ 声の色々な特徴を観察(続)



基本周波数(F0)

第二フォルマント周波数(F2)

第一フォルマント周波数(F1)

4. おわりに
ー 最先端研究への発展

本日のまとめ

- コンピュータで音や声を扱うために重要な「周波数分析」について概説・演習した.
- 本日の内容は古典的な理論が中心ですが, 現在の音響・音声分野におけるすべての技術の基礎です.
- 高校等で学ぶ(学んだ)物理や数学がたくさん登場します.
  - 英語のリスニング, 生物(聴覚や発話)なども関連しています.