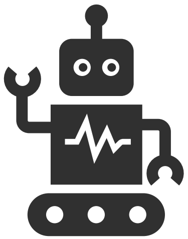
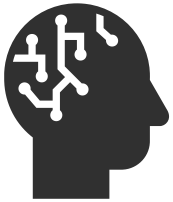
# 機械学習のための音声データの扱い

# 音声データとAI (1/2)

- **音声データを処理する能力**を持つAIの研究・開発が近年多くなされている.
  - 利便性の向上
  - 業務の生産性の向上
  - 他の技術と組み合わさることができる
- 活用事例：
  - スマートスピーカー
  - 音声アシスタント
  - 会議などで使われる自動議事録AI





# 音声データとAI (2/2)

音声認識をタスクとしたデータ分析コンペも多数

- **Kaggle Freesound Audio Tagging 2019**
  - 短い音声データからギターや犬の鳴き声などタグ付けするタスク
- **Kaggle BirdCLEF2021: Processing audio data**
  - 鳥の鳴き声の音声データから，各鳴き声に対応する鳥の種類を推測するタスク

→ **音声データをどうやって扱うか？**

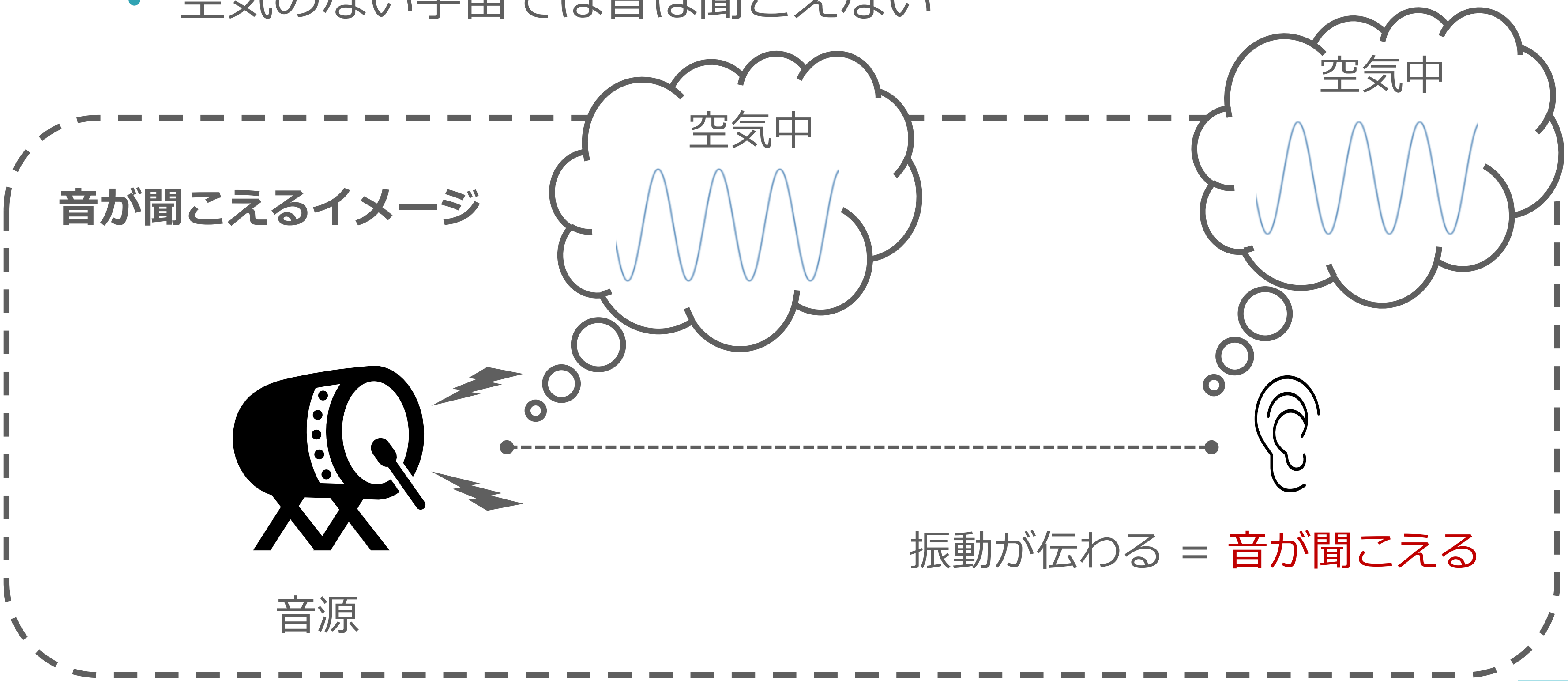
そもそも，

**音声データとはなんだろうか**

# 音が聞こえる仕組み

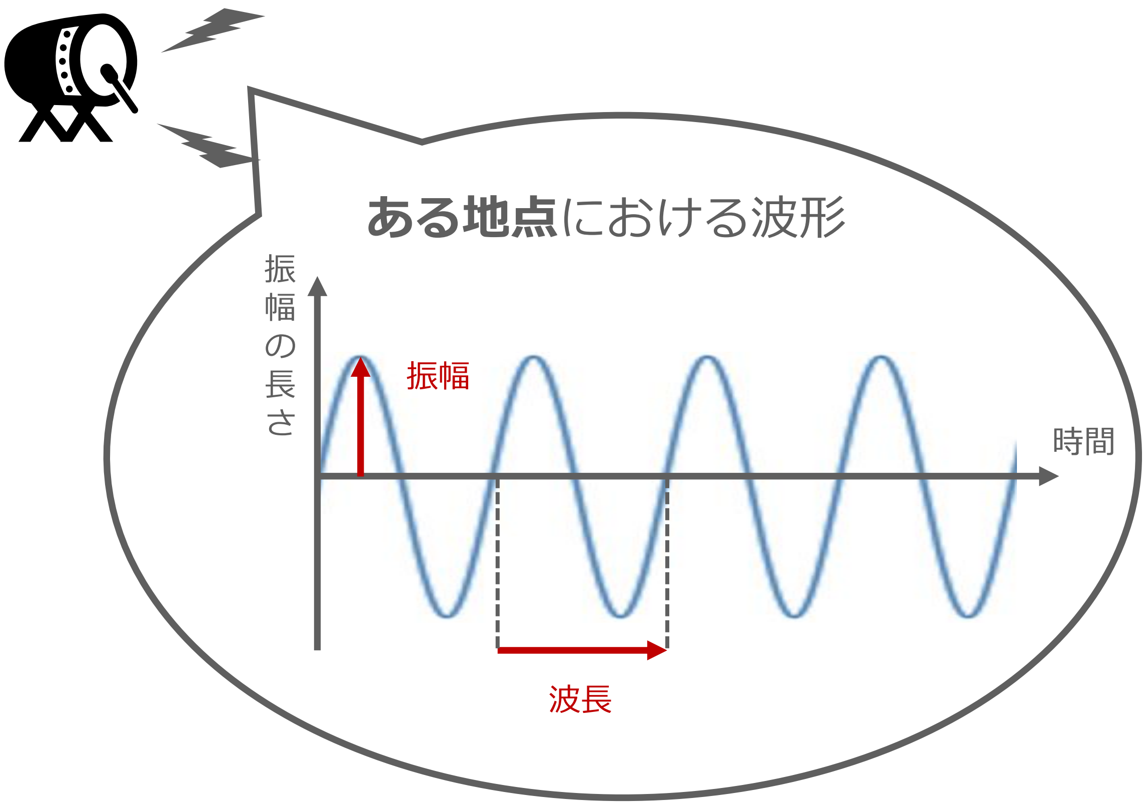
## 音はどのように耳に伝わるか

- 物体の振動による空気の振動
- 空気のない宇宙では音は聞こえない



# 音波

## 空気の振動による**音の波**



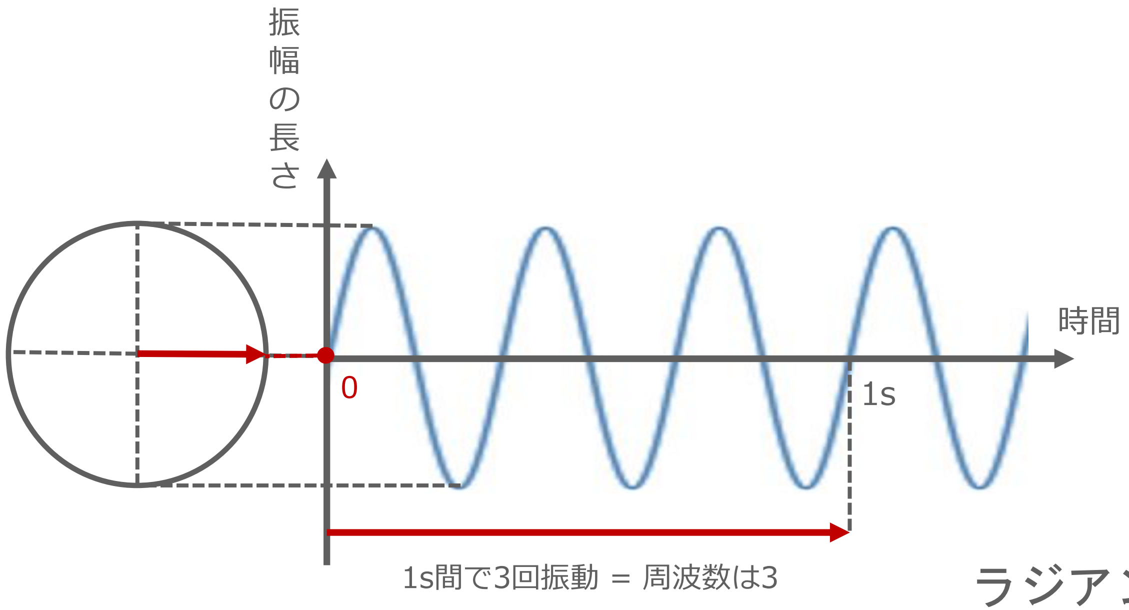

**振幅: 音の大きさ**
**波長 : 音の高さ**

振幅が大きい → 大きい音
波長が大きい → 低い音

**1波長分の時間 = 1周期**

# 周波数と角周波数

- 周波数：一秒あたりの振動数(周期数)
- 角周波数：周波数を回転する角度で表現

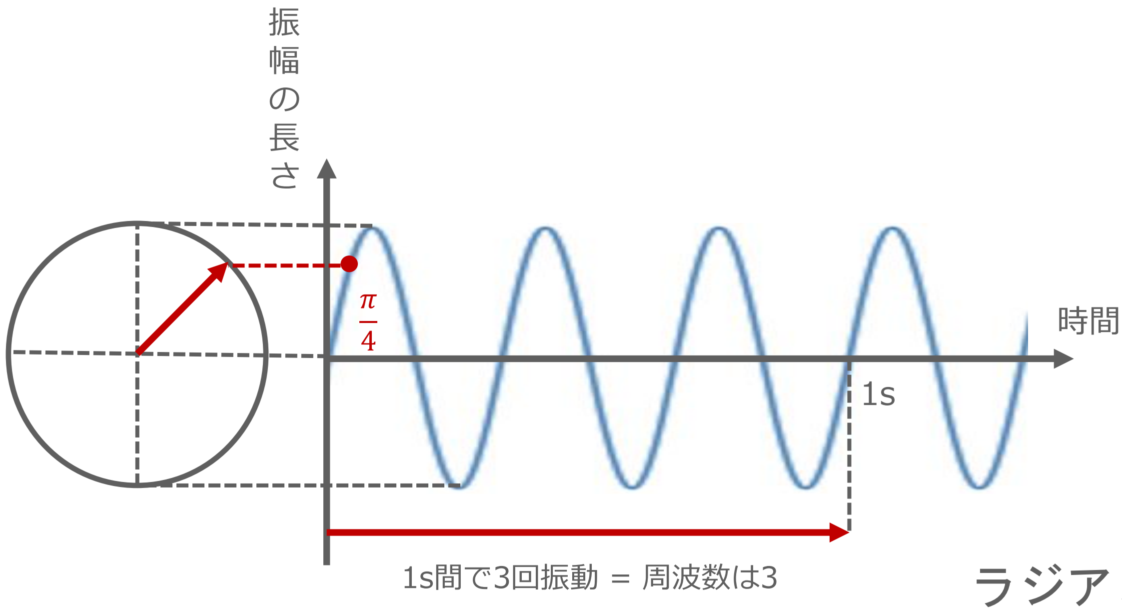



ラジアンの確認

$$\pi = 180°$$
$$\pi/2 = 90°$$
$$\pi/4 = 45°$$

# 周波数と角周波数

- 周波数：一秒あたりの振動数(周期数)
- 角周波数：周波数を回転する角度で表現




ラジアンの確認

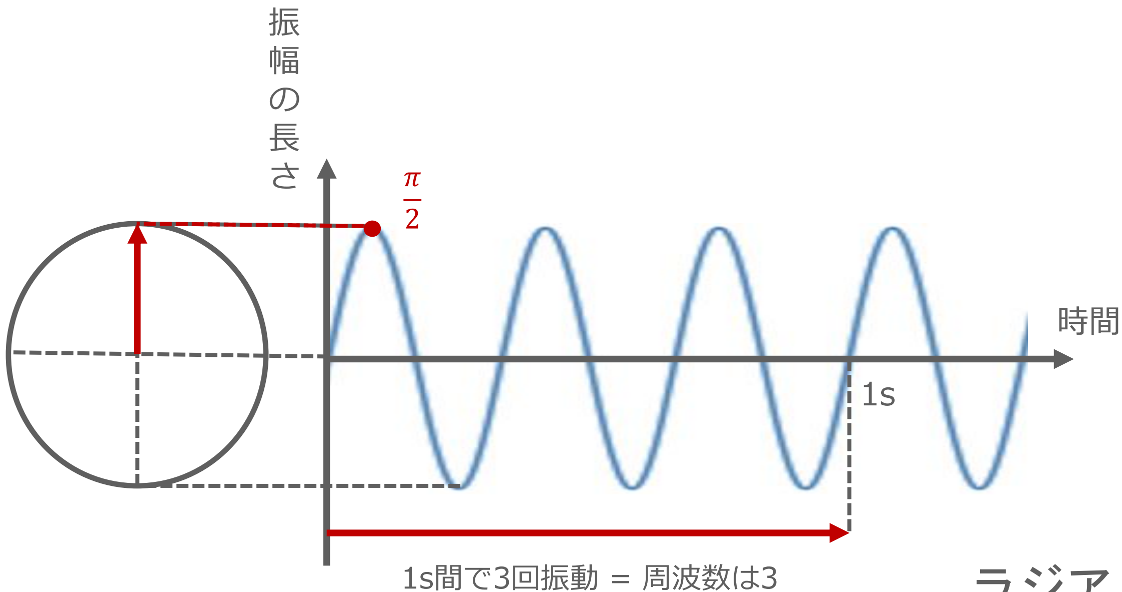
$\pi = 180°$

$\pi/2 = 90°$

$\pi/4 = 45°$

# 周波数と角周波数

- 周波数：一秒あたりの振動数(周期数)
- 角周波数：周波数を回転する角度で表現

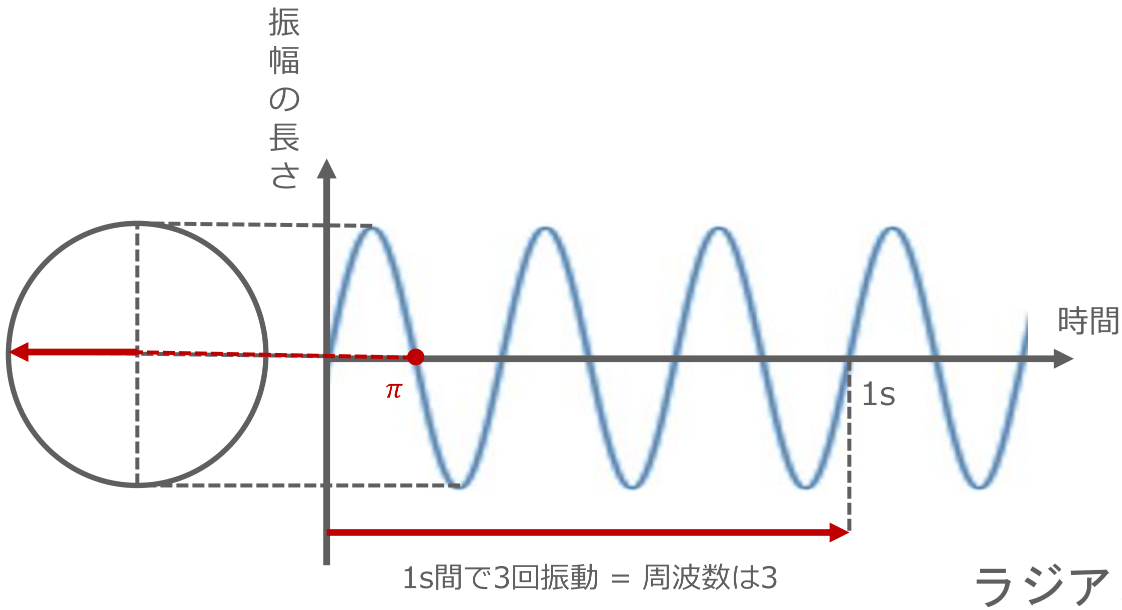



ラジアンの確認

$$\pi = 180°$$
$$\pi/2 = 90°$$
$$\pi/4 = 45°$$

# 周波数と角周波数

- 周波数：一秒あたりの振動数(周期数)
- 角周波数：周波数を回転する角度で表現




ラジアンの確認

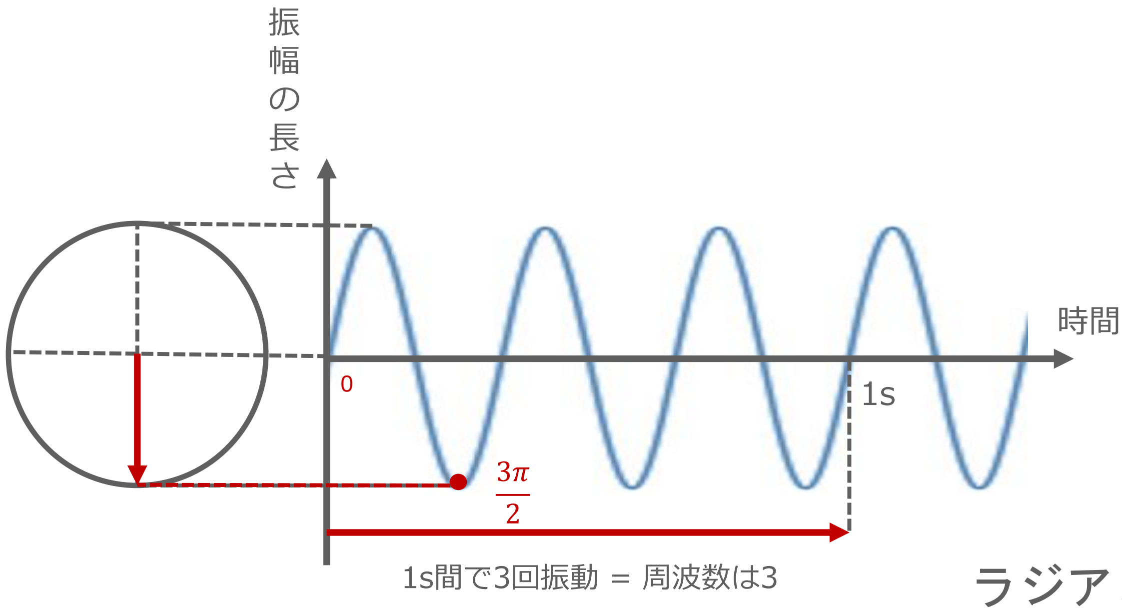
$\pi = 180°$

$\pi/2 = 90°$

$\pi/4 = 45°$

# 周波数と角周波数

- 周波数：一秒あたりの振動数(周期数)
- 角周波数：周波数を回転する角度で表現




ラジアンの確認

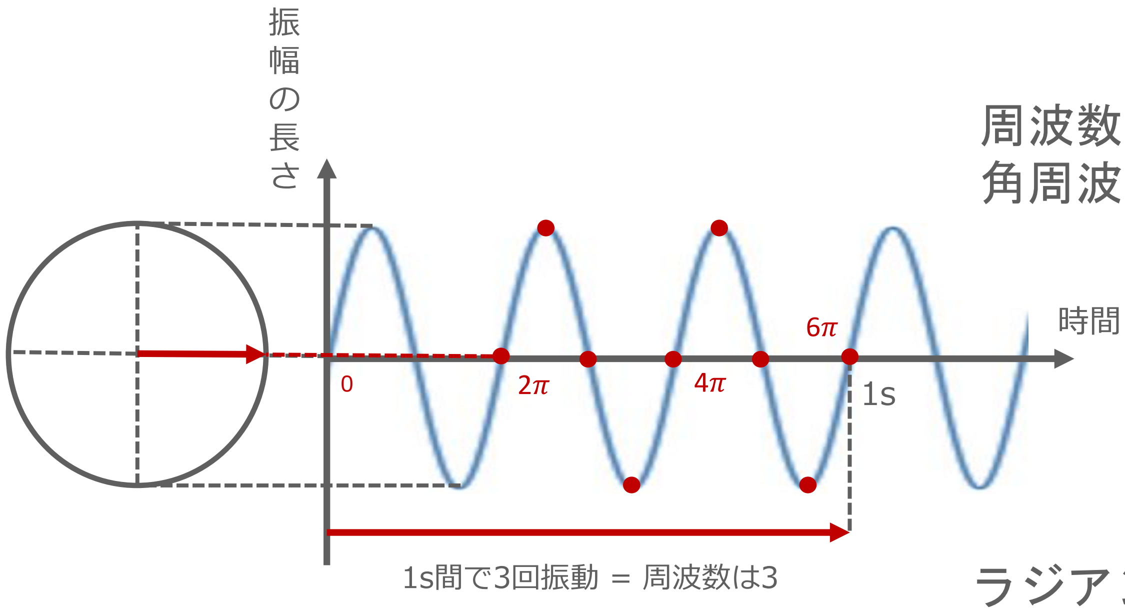
$\pi = 180°$

$\pi/2 = 90°$

$\pi/4 = 45°$

# 周波数と角周波数

- 周波数：一秒あたりの振動数(周期数)
- 角周波数：周波数を回転する角度で表現




周波数：3,
角周波数：$6\pi$

ラジアンの確認
$$\pi = 180°$$
$$\pi/2 = 90°$$
$$\pi/4 = 45°$$

# 音波と機械学習

波形
(非周期的でも可)

波形処理

- 標本化
- 量子化
- フーリエ変換

自然言語処理

e.g. スマートスピーカなど

分類タスク

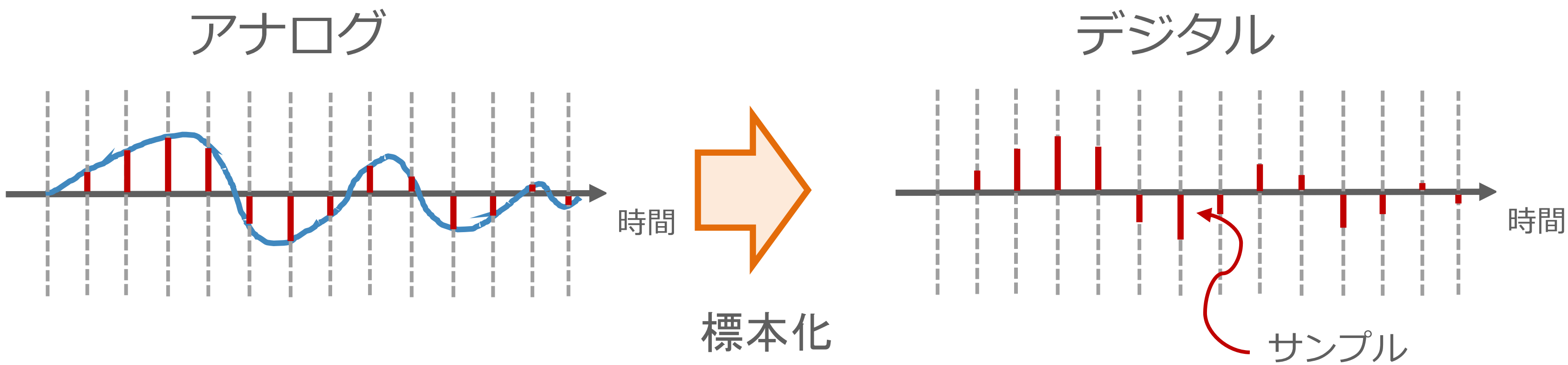
e.g. 鳥の種類予測など

未来予測

音データ以外への応用
e.g. 売り上げなど

# 波形の扱い(1/2)

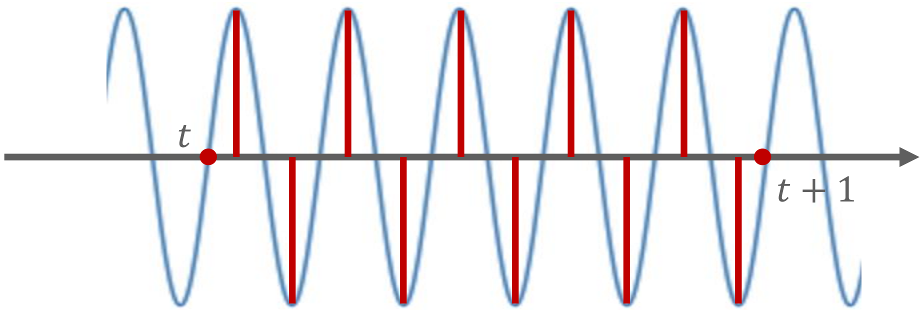
**標本化**：連続時間信号を離散時間信号に変換




**量子化**：等分した振幅にサンプルの振幅を合わせる

# 波形の扱い (2/2)

- **サンプリング周波数：**
  - １秒間で処理することができるサンプルの個数
- **サンプリングの法則**
  - 周波数h[kHz]の離散時間信号を測るには，最低2h[kHz]のサンプリング周波数が必要




周波数6 → サンプリング周波数12が必要

# フーリエ変換とは (1/3)

**目的**：波形を機械学習の入力とするために行う

(標本化，量子化と併用)

**前提**：あらゆる波形(周期的・非周期的)は,

正弦波・余弦波を用いて表現できる

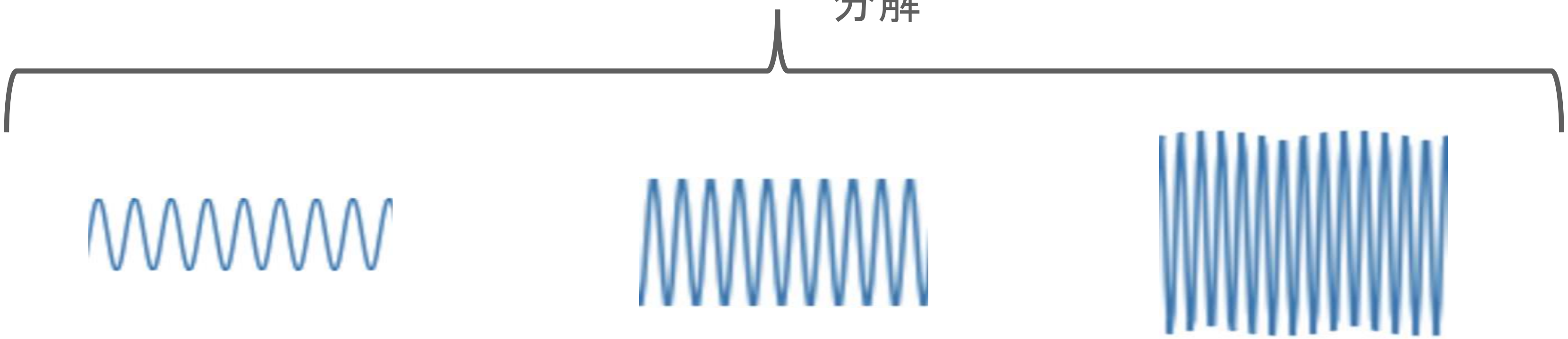
**振幅と周波数が分かれば，波の特性がわかる！**

振幅$h$，角周波数$\omega$の正弦波：$h sin\omega$

振幅$h$，角周波数$\omega$の余弦波：$h cos\omega$

# フーリエ変換とは (2/3)


分解



$sinx$
振幅1,角周波数$\pi$

$2sin3x$
振幅2,角周波数$3\pi$

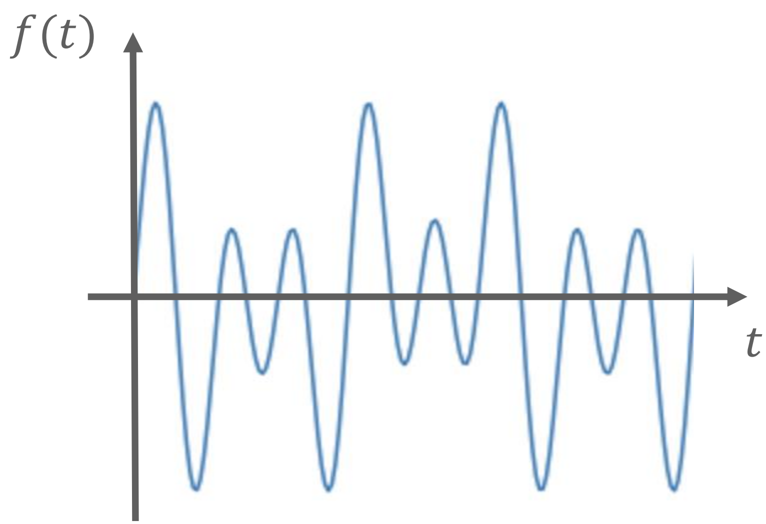
$3sin5x$
振幅3,角周波数$5\pi$

ある波形を構成する波の振幅と周波数を機械学習の入力とする

# フーリエ変換とは (3/3)

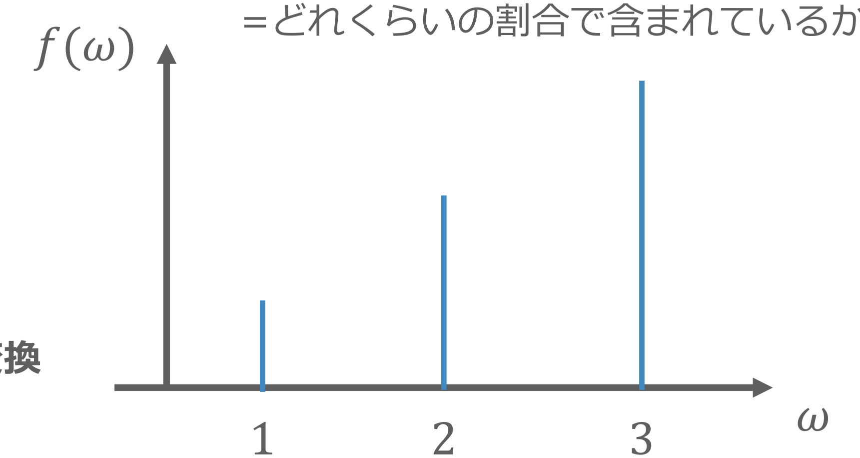
**定義**：ある波形$f(t)$から振幅・角周波数を表す

関数$F(\omega)$に変換する作業




フーリエ変換

縦軸は各波の振幅
＝どれくらいの割合で含まれているか




各波の角周波数

この図をスペクトルと呼ぶ

# フーリエ変換の概算(補足)

- 全ての波形は正弦波・余弦波で表せる

$$f(x) = a + \sum_{n=1}^{\infty} a_n cosnx + b_n sinnx$$

(マクローリン展開より)

- オイラーの公式:$e^{i\theta} = cos\theta + isin\theta$などを用いて,

## フーリエ変換の公式

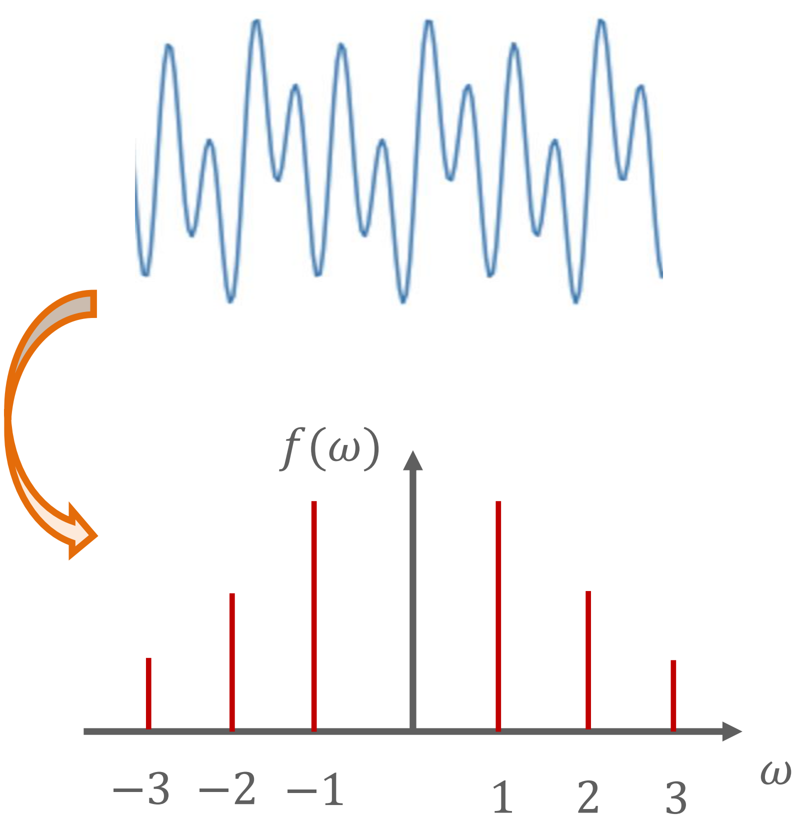
※$f(t)$はある条件を満たす

$$\underbrace{f(x)}_{\text{波形}} = \underbrace{\int_{-\infty}^{\infty}}_{\text{波の合成}} [\underbrace{\int_{-\infty}^{\infty} f(t)e^{-i\omega t}dt}_{\text{ある波の振幅}}] \underbrace{e^{i\omega x}}_{\text{ある周波数の波}} d\omega$$

ある波の振幅：$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t}dt$

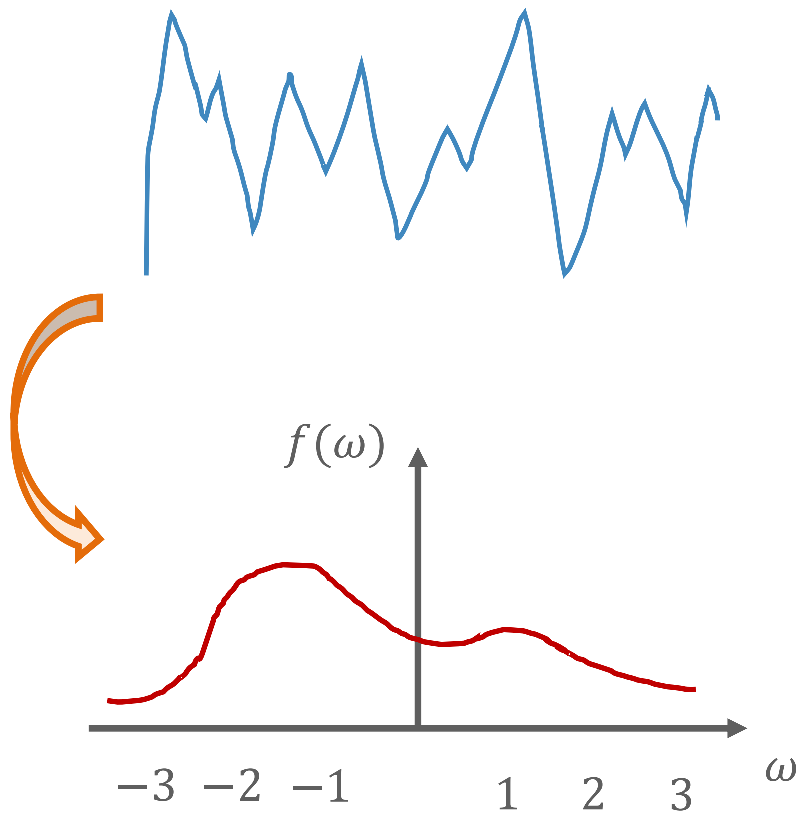
# スペクトル

## 周期的波形




離散スペクトル

## 非周期的波形



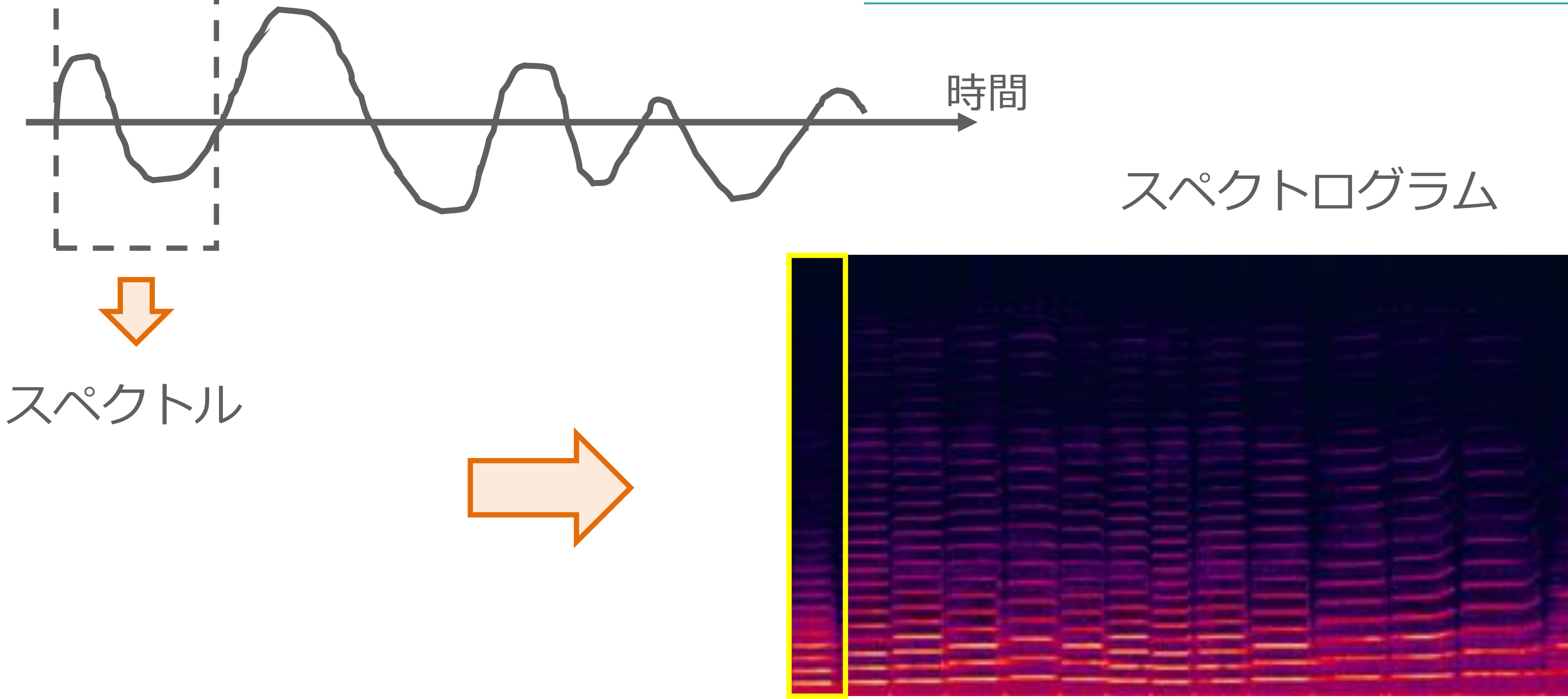

連続スペクトル

# スペクトログラム

目的：現実的である非周期音声データの分析

窓関数：波形を特定の時間区間(窓)で区切る

https://ja.wikipedia.org/wiki/%E3%82%B9%E3%83%9A%E3%82%AF%E3%83%88%E3%83%AD%E3%82%B0%E3%83%A9%E3%83%A0



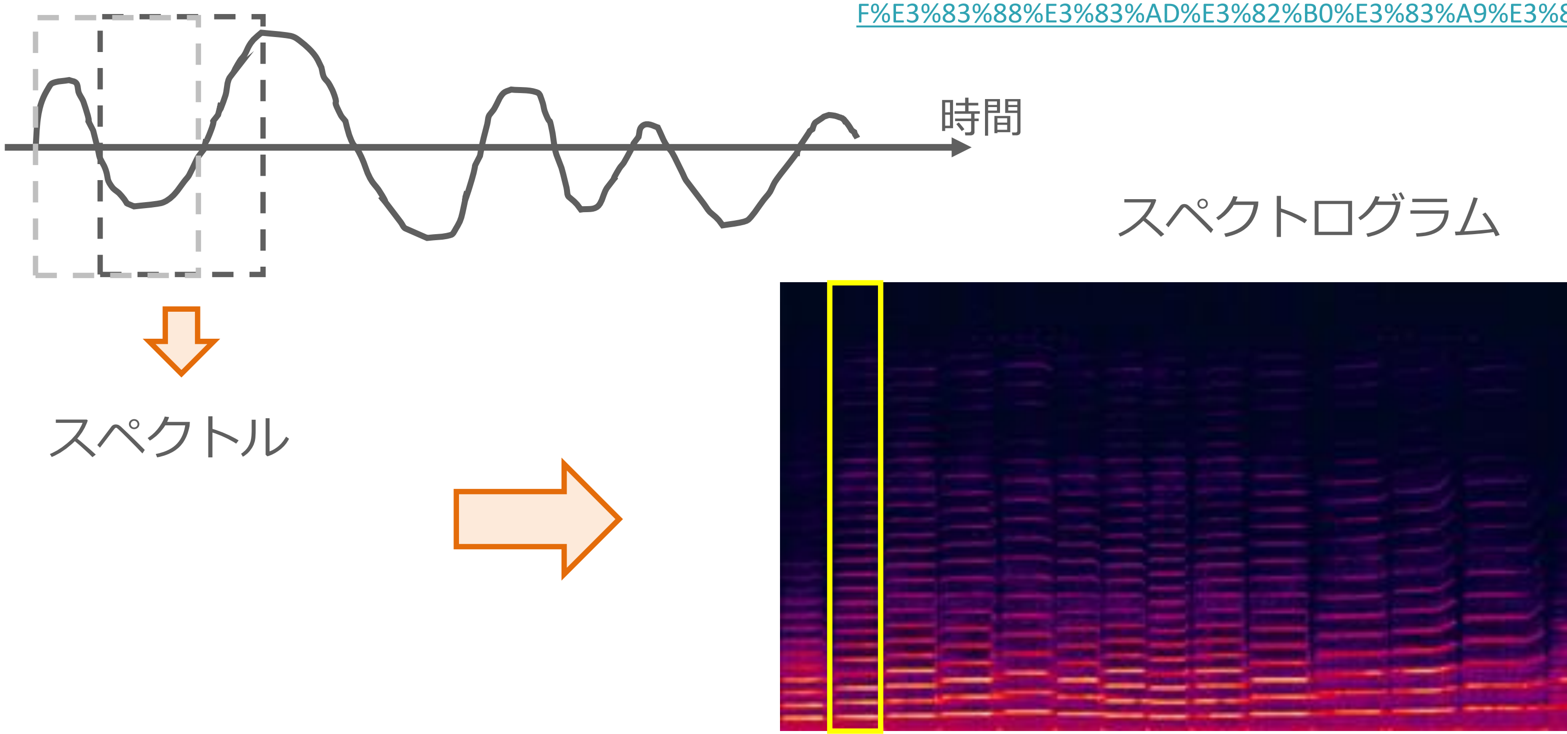

横軸：時間，縦軸：周波数，輝度：振幅

# スペクトログラム

目的：現実的である非周期音声データの分析

窓関数：波形を特定の時間区間(窓)で区切る

https://ja.wikipedia.org/wiki/%E3%82%B9%E3%83%9A%E3%82%AF%E3%83%88%E3%83%AD%E3%82%B0%E3%83%A9%E3%83%A0



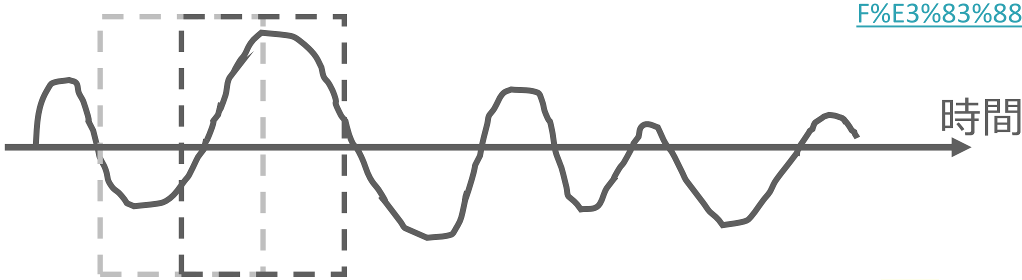

横軸：時間，縦軸：周波数，輝度：振幅

# スペクトログラム

目的：現実的である非周期音声データの分析

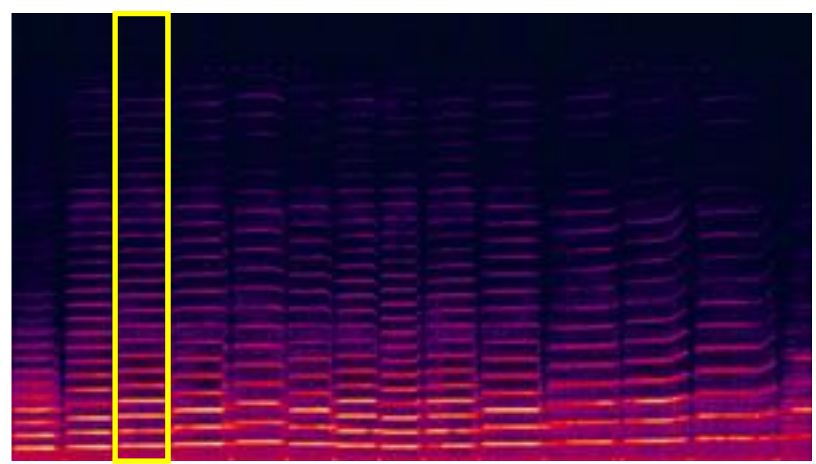
波形を特定の時間区間(窓)で区切る

https://ja.wikipedia.org/wiki/%E3%82%B9%E3%83%9A%E3%82%AF%E3%83%88%E3%83%AD%E3%82%B0%E3%83%A9%E3%83%A0




スペクトル

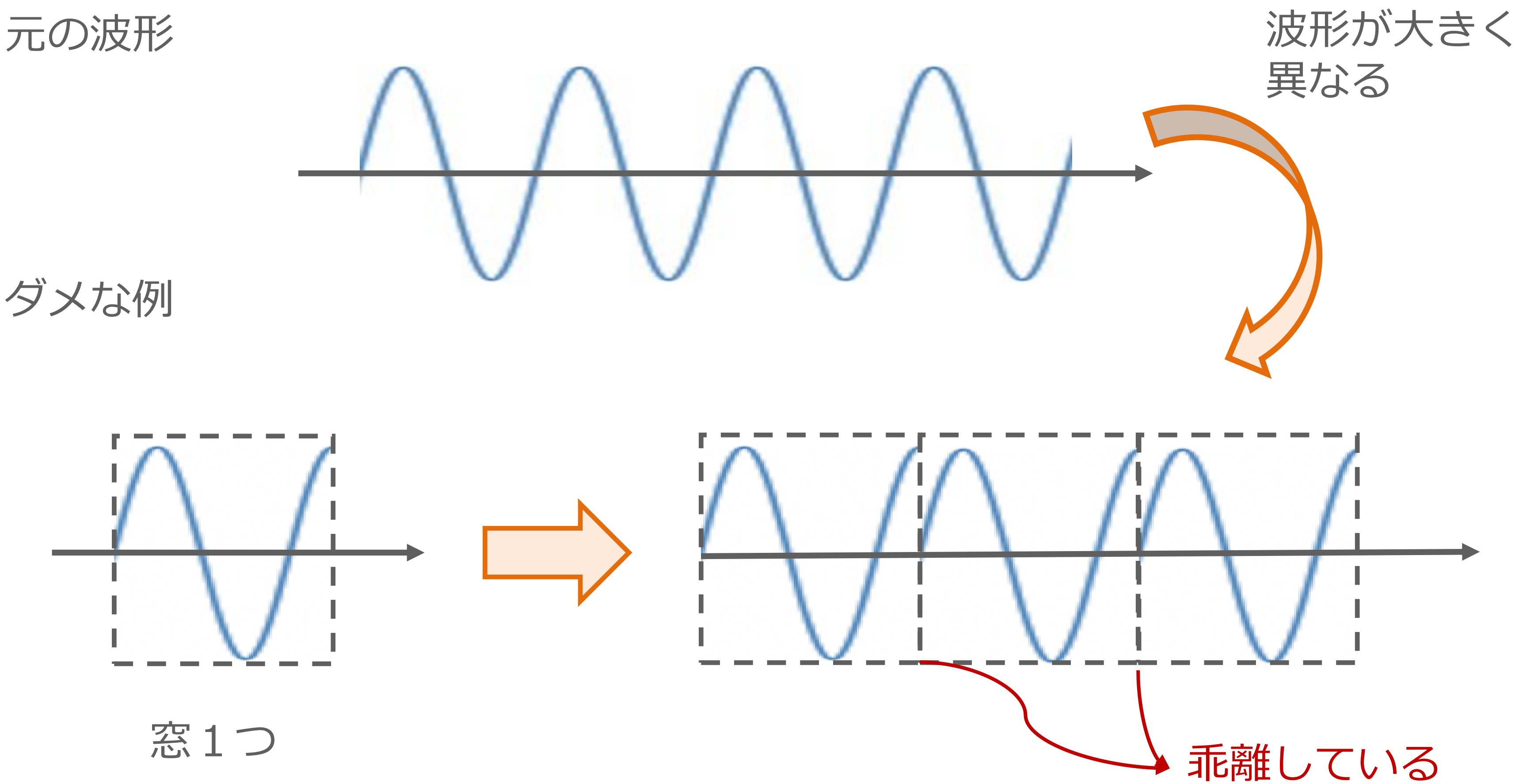
スペクトログラム



横軸：時間，縦軸：周波数，輝度：振幅

# 窓関数 (1/2)

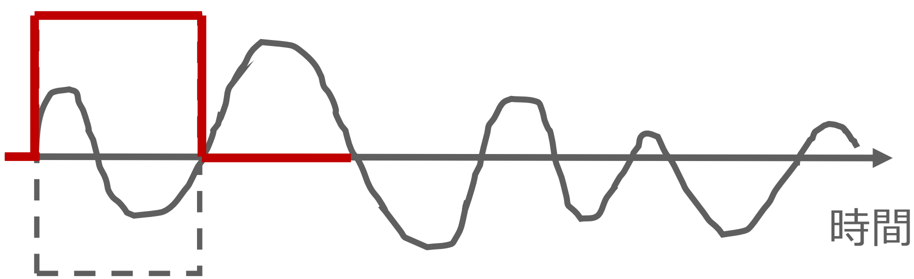
- 窓の大きさが問題となる



# 窓関数 (2/2)

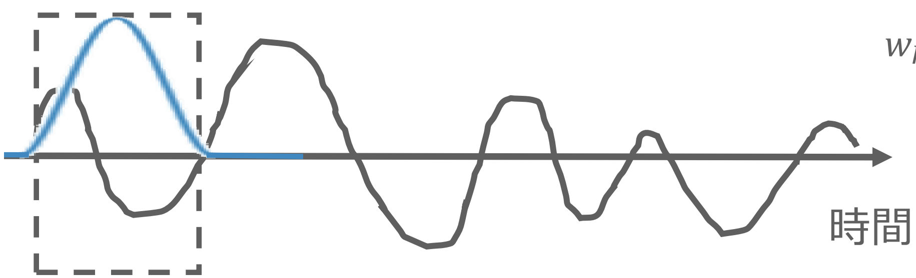
- 窓のつなぎ目を**滑らか**にするために区間ごとの波形関数にかける関数

- 普通の窓：矩形窓




$$w_r[n] = \begin{cases} 1\ (n = 1,2,\ldots,N-1) \\ 0\ (otherwise) \end{cases}$$

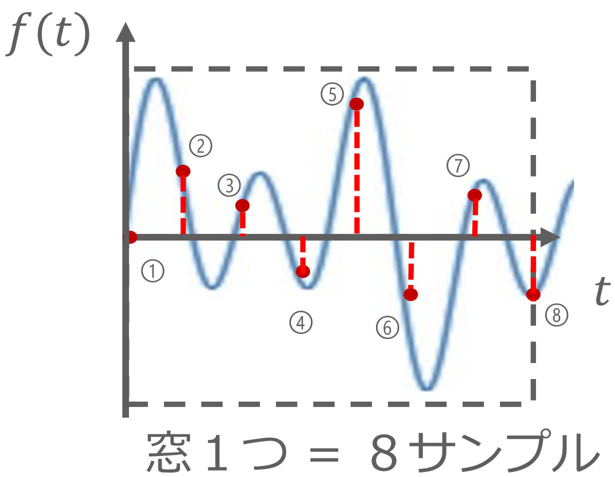
- **ハミング窓**

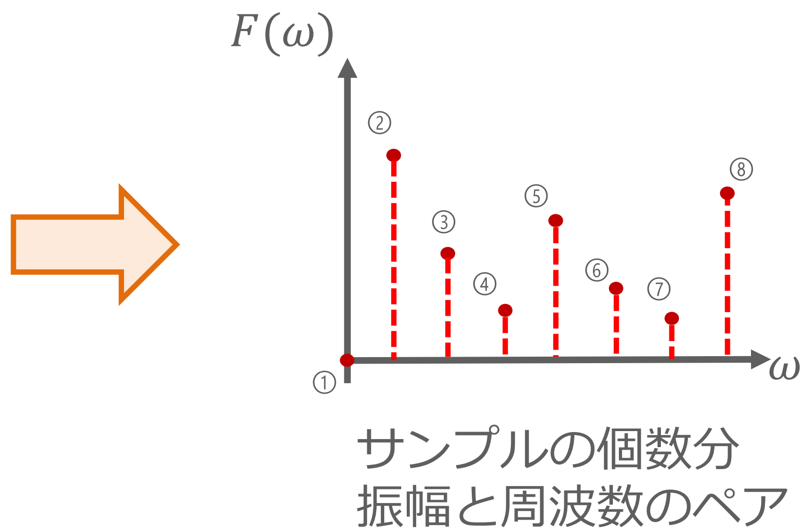



$$w_h[n] = \begin{cases} 0.54 - 0.46cos\dfrac{2\pi n}{N}\ (n = 1,2,\ldots,N-1) \\ 0\ (otherwise) \end{cases}$$

# DFTとFFTにおけるサンプル数

## DFT：離散フーリエ変換




## FFT：高速フーリエ変換

窓のサンプルのうち，偶数番目と奇数番目を別々に測定

窓に含まれるサンプルN個 (Nは２の冪乗：8,16,32,⋯,1024,2048, ⋯)

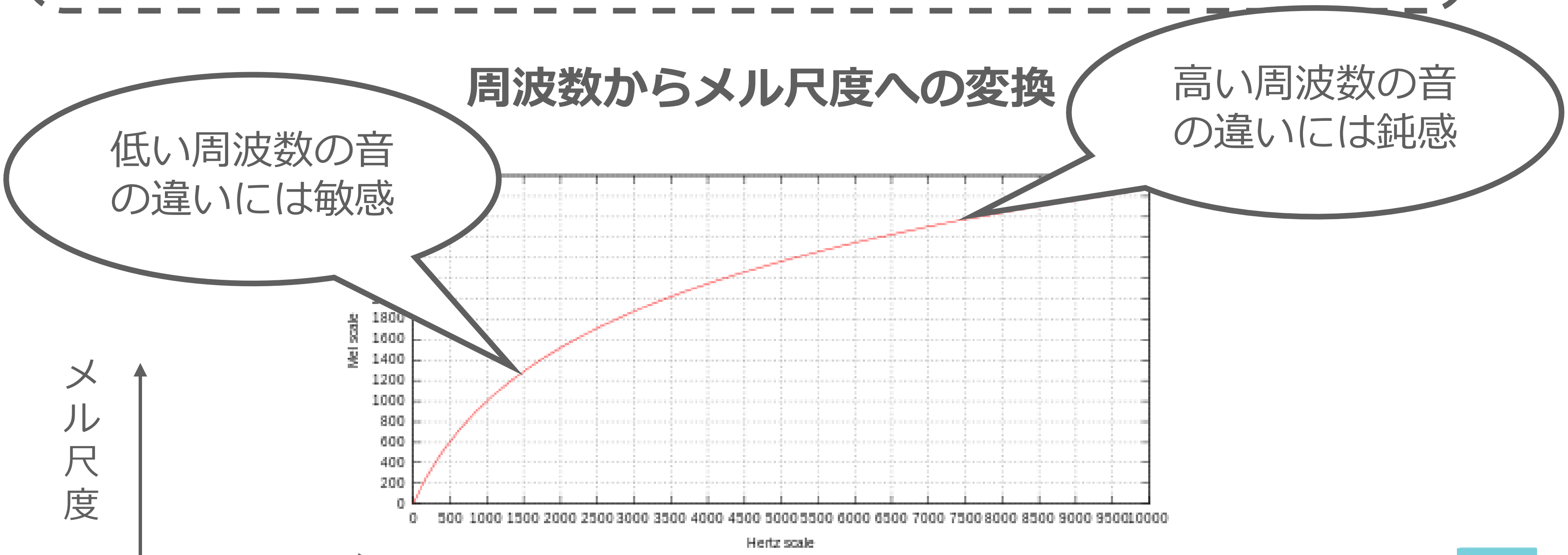
→ N/2個のサンプルを測定　→　**高速化**

# その他の技術 (1/2)

- **メル尺度**：人間の聴覚に基づいた尺度

周波数の低い音に対して敏感で、周波数の高い音に対して鈍感であるという性質がある

**周波数からメル尺度への変換**




https://ja.wikipedia.org/wiki/%E3%83%A1%E3%83%AB%E5%B0%BA%E5%BA%A6

# その他の技術 (2/2)

- **逆フーリエ変換**
  - 振幅・周波数から元の波形を構築する作業
- **ケプストラム**
  - フーリエ変換したものの絶対値の対数を逆フーリエ変換して得られるもの
  - 音声認識の特徴料として利用される