

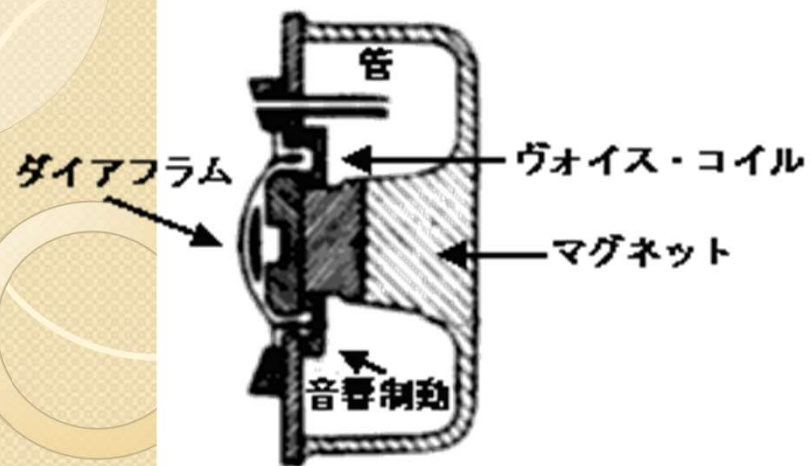


実世界情報処理

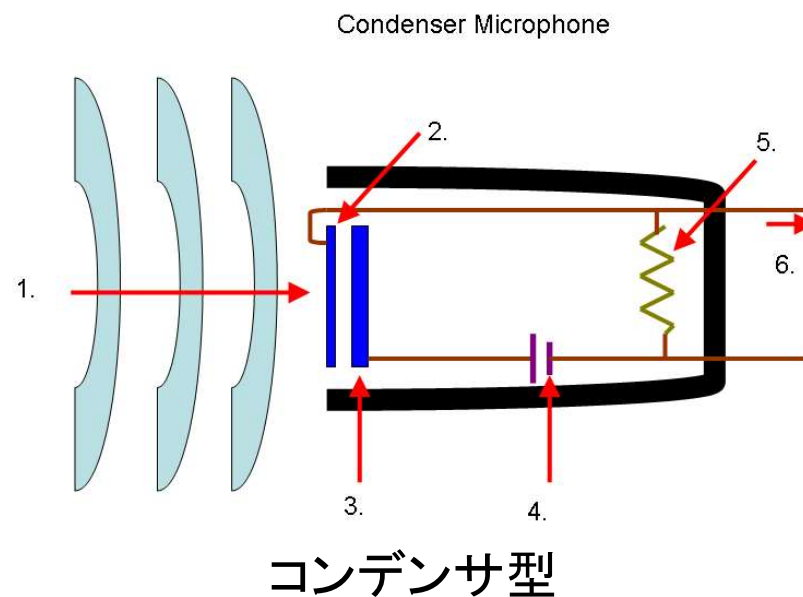
(音声インタフェース)

マイクロフォン(Microphone)

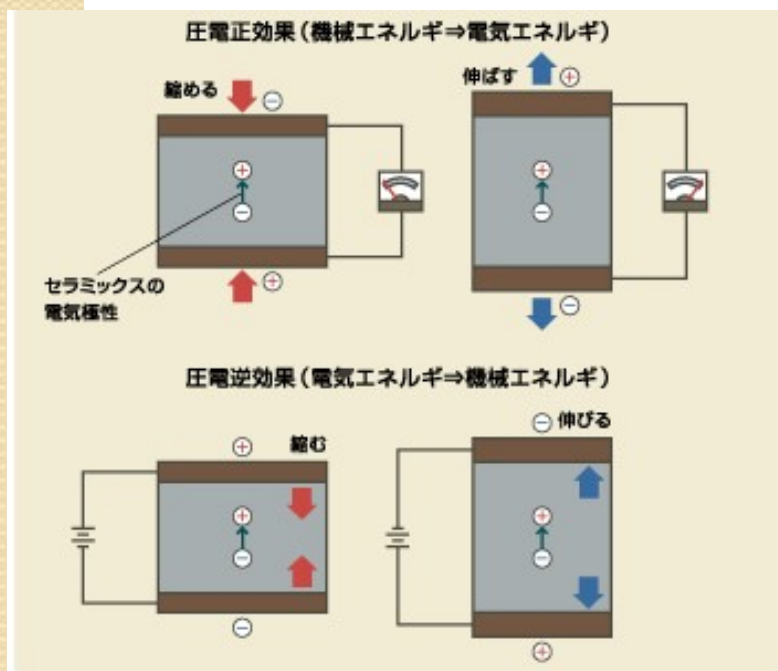
- 音を電気信号に変換するセンサの一種
 - 空気の振動がマイクの中の素子を振動させ、それによって電気信号が変化する。
- 種類
 - ムービングコイル型
 - コンデンサ型
 - 圧電マイク
 - レーザマイク



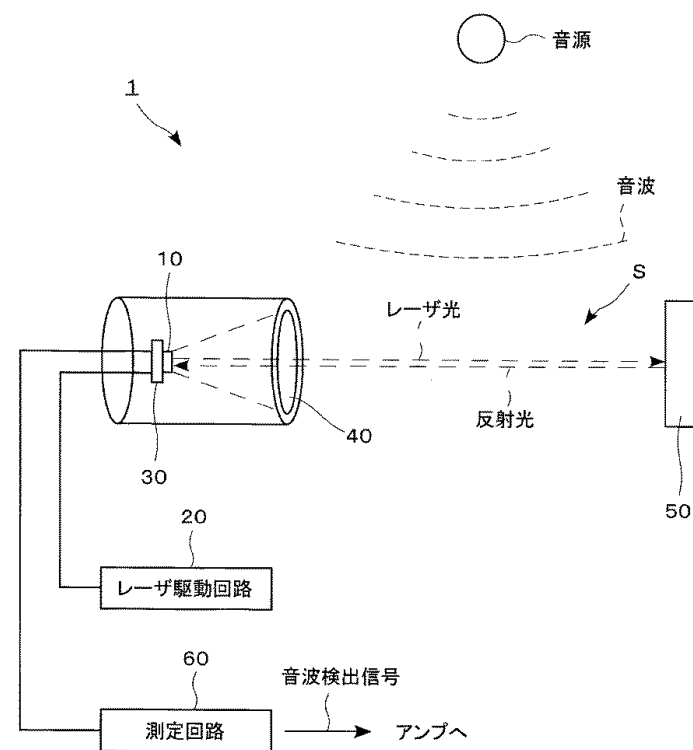
ムービングコイル型



コンデンサ型

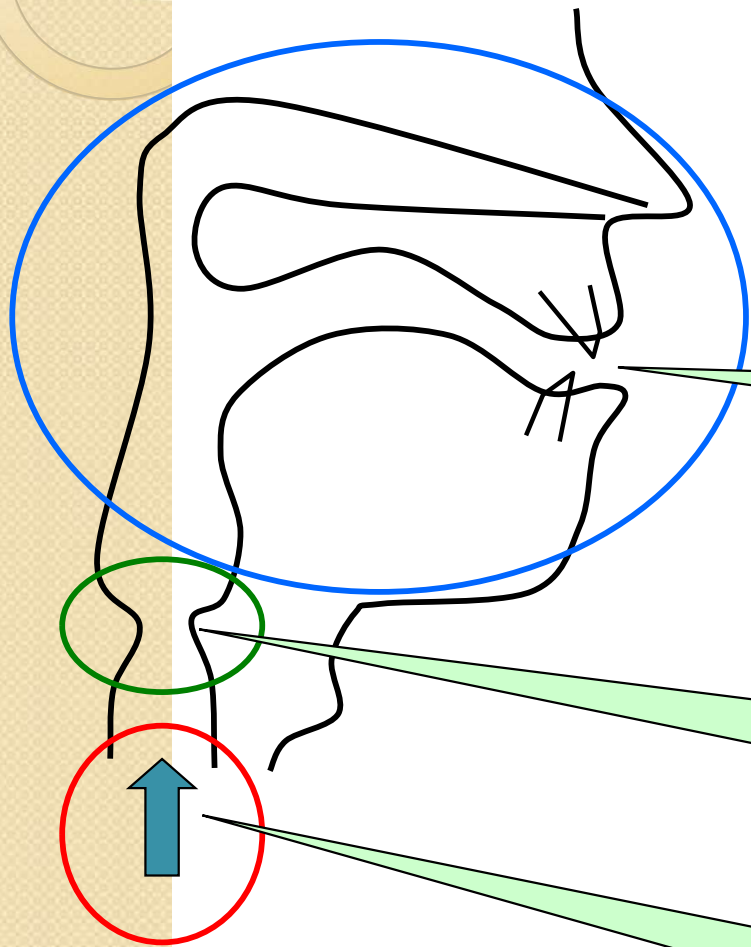


圧電マイク



レーザマイク

音声の成り立ち



音素(phoneme)

音の発生の仕方

声帯振動、摩擦、破裂



/a/, /i/, /u/, ...

/k/, /s/, /t/, ...

声道、口、鼻腔などの特性

アクセント(accent)

声帯振動の周期

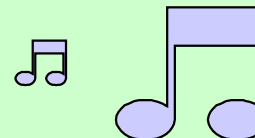
→ 音程



わたしはにほんじんです

息の強弱

→ 強勢



I'm an American

用語（音）

- 単音/音（phone）

実際に発話された音。

[r], [l], [ŋ], [n] などのように [] でくくる。

- 音素（phoneme）

人間の認識する音。

/r/, /N/ などのように / / でくくる。

- [r], [l] → 日本人にrとlの区別はつかないのでどちらも/r/
- [n] → 文脈によって「なにぬねの」の/n/、「ん」の/N/

- 音節（syllable）

母音のように安定した音を中心とした音素の集まり。

子音×(0～3) + 母音 + 子音×(0～3)

- 英語 ・ ・ ・ a も strength も1音節。
- 日本語 ・ ・ ・ 1子音+1母音。音節の長さがほぼ一定。

音声認識とは

- 入力された音声から、音節・単語を判断する
 - 単音・音素単位での認識は無理
 - ・ 文脈によって音と音素の対応が変わる
 - ・ 曖昧な音・発音ミスを、文脈で訂正

具体的な処理

- 音列 $Y = (y_n, y_{n-1}, y_{n-2}, \dots)$ が与えられたとき、条件付確率 $P(W | Y)$ を最大にするような単語列 $W = (w_n, w_{n-1}, w_{n-2}, \dots)$ を探索・推定
- 実際には、 Y が変数なのに $P(W | Y)$ を求めるのは難しいので、ベイズの定理を使って、以下の式で計算

$$P(W | Y) = P(W)P(Y | W) / P(Y)$$

$P(Y | W)$... 音響モデルを使って計算

$P(W)$... 言語モデルを使って計算

$P(Y)$... W には無関係

単語wの発音はy

単語w0の後には単語w1
が来ることが多い

音響モデルと言語モデル

$$P(W | Y) = P(W)P(Y | W) / P(Y)$$

- 音響モデル

- 単語列 W と音声列 Y の合致度 $P(Y | W)$ を求める
 - 特徴量解析 → ケプストラム分析
 - 確率・統計的モデル → 隠れマルコフモデル (HMM)

- 言語モデル

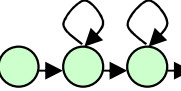
- 単語の発声確率 $P(W)$ を言語の特徴から計算

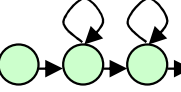
隠れマルコフモデル (HMM)

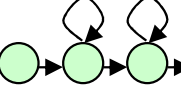
音声列 Y を入力 y_1, y_2, y_3, \dots

各 y_i は、ケプストラム等の音響的特徴量の実数ベクトル、もしくはそれをベクトル量子化したもの

HMMデータベース

Λ_0 : 単語 w_0 の音を出ししやすいHMM

Λ_1 : 単語 w_1 の音を出ししやすいHMM

Λ_2 : 単語 w_2 の音を出ししやすいHMM

\vdots

Λ_M : 単語 w_M の音を出ししやすいHMM

- 音声列 Y を最も出力しやすいHMMを探索

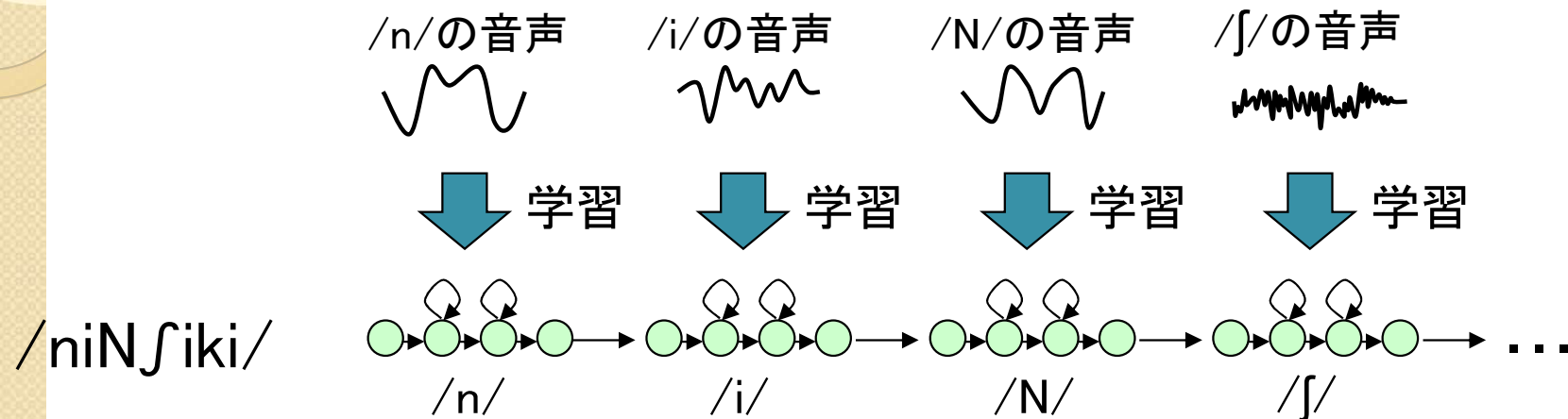
$$x = \arg \max_{i \in M} [P(Y | \Lambda_i)]$$

- そのHMMに対応する単語 w_x を出力

単語列 W を出力 w_1, w_2, w_3, \dots

HMMデータベースの作成

1. 音素単位のHMMを作って連結(mono-phoneモデル)



2. 前後の音素も考慮 (bi-phone, tri-phoneモデル)

後ろの音素を考慮 $/n+i/ \rightarrow /i+N/ \rightarrow /N+ʃ/ \rightarrow /ʃ+i/ \rightarrow \dots$

前の音素を考慮 $/n/ \rightarrow /i-n/ \rightarrow /N-i/ \rightarrow /ʃ-N/ \rightarrow \dots$

前後の音素を考慮 $/n+i/ \rightarrow /i-n+N/ \rightarrow /N-i+ʃ/ \rightarrow /ʃ-N+i/ \rightarrow \dots$