東京大学工学部 4年生 夏学期 [03-501130]

応用音響学
第1回 (4/08)

亀岡弘和

東京大学大学院情報理工学系研究科
システム情報学専攻
kameoka@hil.t.u-tokyo.ac.jp

講義スケジュール

前半（亀岡担当）

- 4/08: 第1回
- 4/15: 第2回
- 4/22: 第3回
- 4/29: 休日
- 5/01(木): 第4回
- 5/06: 休日
- 5/13: 第5回
- 5/20: 第6回
- 5/27: 第7回

後半（牧野担当）

- 6/03: 第8回
- 6/10: 第9回
- 6/17: 第10回
- 6/24: 第11回
- 7/01: 第12回
- 7/08: 第13回
- 7/15: 第14回
- 7/22: 学期末試験

講義項目

【講義前半】（亀岡担当）

- 統計的信号処理の基礎
- 音声分析・符号化 の基礎
- 音声認識・合成の基礎
- 音声強調の基礎
- 多チャンネル音響信号処理の基礎
- 音声音響信号処理応用

【講義後半】（牧野担当）

- ヘルムホルツ共振と電気回路とのアナロジー
- 共振のQ値
- 音響インピーダンスと音の放射
- 位相制御による音波の収束
- 固体中の音響

講義内容（キーワード）

【講義前半】（亀岡担当）

- 音声音響信号処理を題材として統計的信号処理や確率モデルの理解を深める。
- 以下のようなキーワードの概念と関連事項を理解することを目標とする。「確率モデル」,「最尤推定量，最小平均二乗誤差推定量」,「ベイズ統計学」,「音声生成の仕組み(調音)」,「音声波形とスペクトル」,「線形予測分析」,「音声分析合成」,「偏自己相関分析」,「k-meansクラスタリング」,「混合正規分布モデル(GMM)」,「動的時間伸縮マッチング」,「隠れマルコフモデル(HMM)」,「音声合成」,「テキスト音声合成」,「マイクロフォンアレイ」,「非負値行列因子分解」

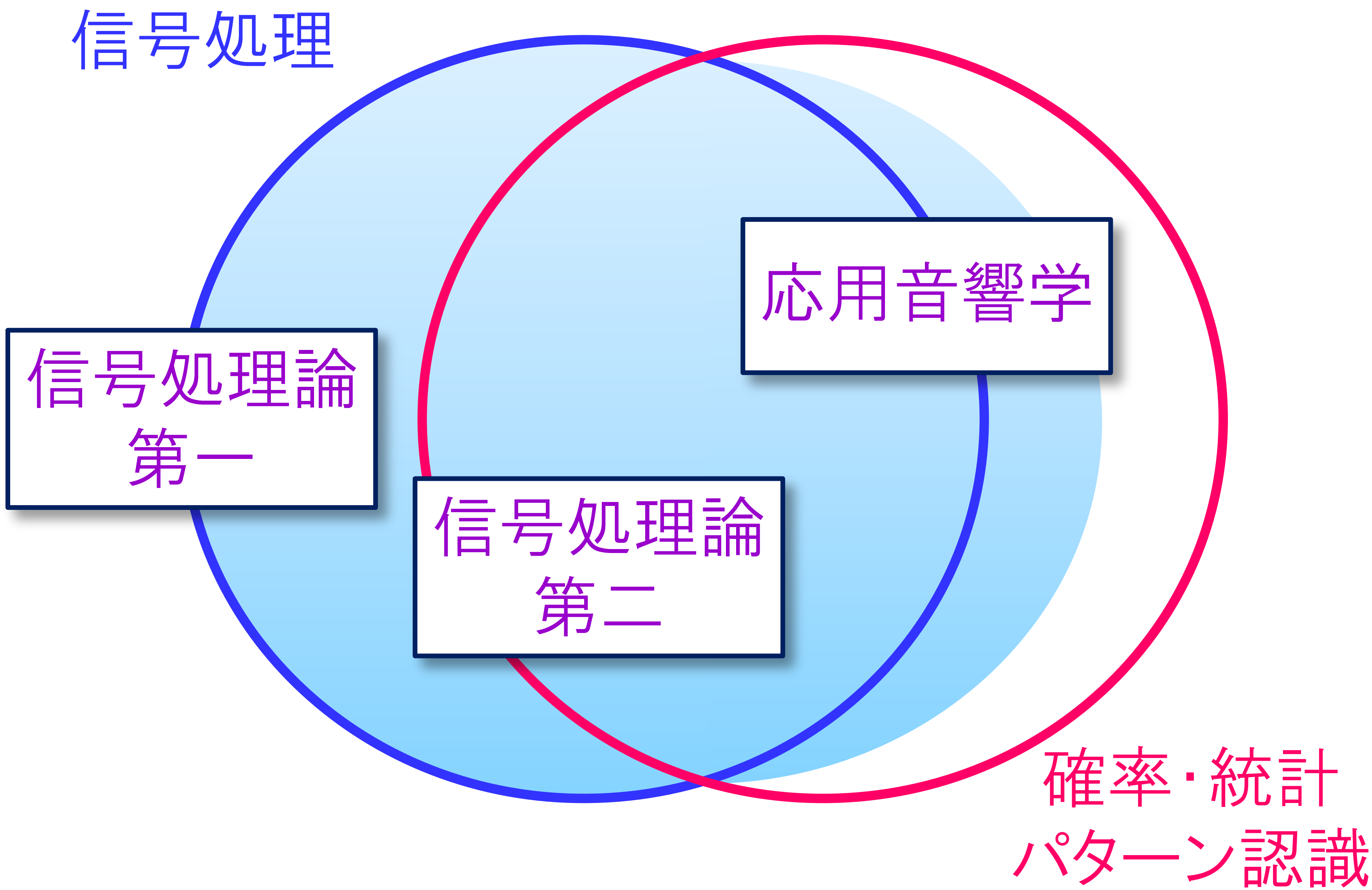
【講義後半】（牧野担当）

- 音響における波動現象を理解し，それとのアナロジーを元に，光学，回路学，電磁気学等，他の分野における同様の波動現象の理解も深める。
- 以下のようなキーワードの概念と関連事項を理解することを目標とする。「共振・共鳴」,「Q値」,「時定数」,「音響インピーダンス」,「放射インピーダンス」,「呼吸球」,「エクスポネンシャルホーン」

講義前半の概要

- 狙いと性格
  - 「信号処理論第一、第二」に続く「信号処理論第三」と捉えても良い
  - 統計的信号処理、非定常時系列モデル、ベイズ統計学、学習アルゴリズム
  - 音声音響信号処理応用
- 対象領域
  - 信号推定、音声分析、音声符号化
  - 音声認識、音声合成
  - その他の音響信号処理
- 講義重点
  - 統計的信号処理の基礎
  - 基本アルゴリズムの理解
- 前提知識
  - 線形系理論とフーリエ解析
  - 確率統計学の基礎 (分布、推定)

講義前半の内容



講義前半の内容

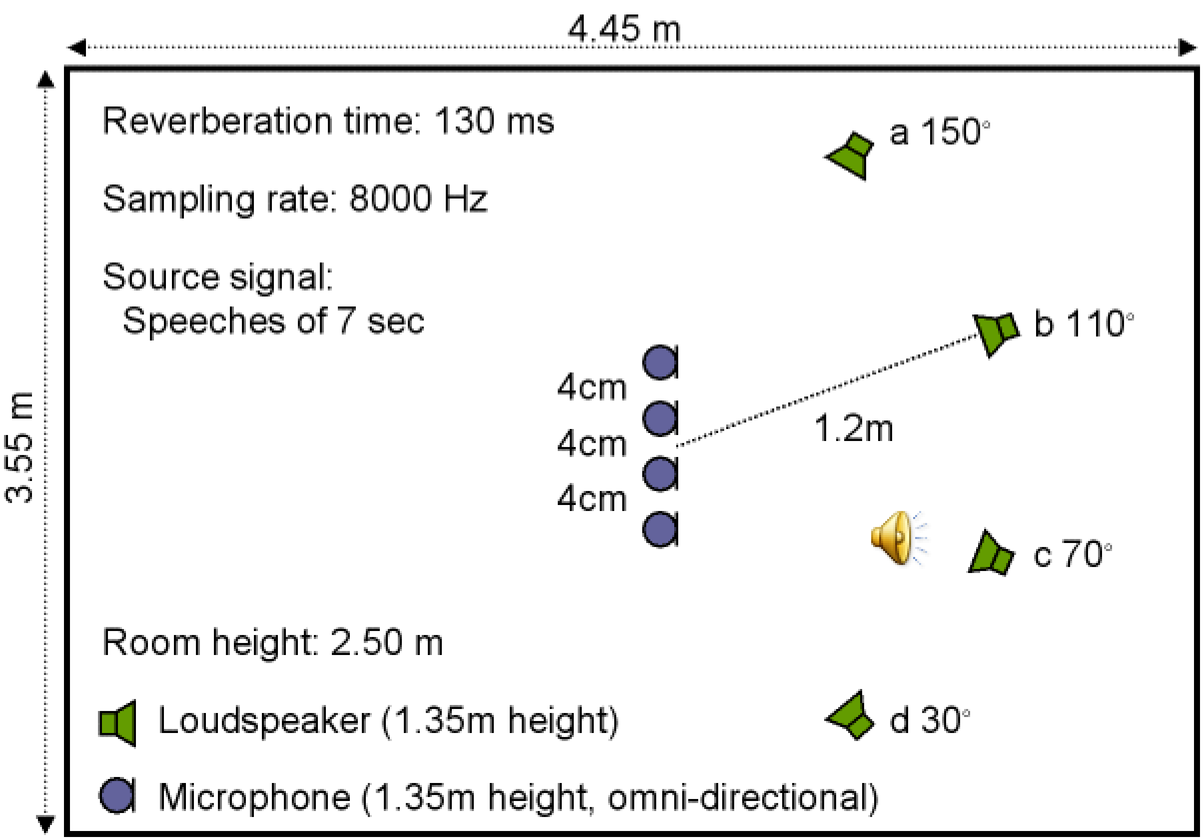
- 信号推定理論
  - Wienerフィルタ、Kalmanフィルタ
- 短時間スペクトル分析
  - サンプリング定理、量子化雑音
  - 高域強調、窓関数
  - 短時間自己相関関数
  - 短時間スペクトル解析、ピッチ構造
  - 短時間ケプストラム解析
- 全極型モデル
  - 線形モデル、自己回帰モデル
  - 線形予測分析(LPC)
  - 残差信号、ピッチ抽出
  - 偏自己相関分析(PARCOR)
- スペクトル距離
  - 板倉齋藤距離
- クラスタリング解析
  - k-means クラスタリング
  - スカラー量子化とベクトル量子化
  - 混合正規分布とEMアルゴリズム
- 非線形時間伸縮
  - 動的時間伸縮、DPマッチング
- 音声の確率モデル
  - 自己回帰モデル
  - 多次元正規分布、混合正規分布
  - 隠れマルコフモデル(HMM)
- 信号処理応用
  - 音声認識／合成システム
  - 音源分離／音声強調／その他の音響信号処理システム

講義資料と成績評価

- 講義資料
  - http://hil.t.u-tokyo.ac.jp/~kameoka/aa/
- 成績評価
  - 出席点
  - 学期末試験

応用紹介：音声認識

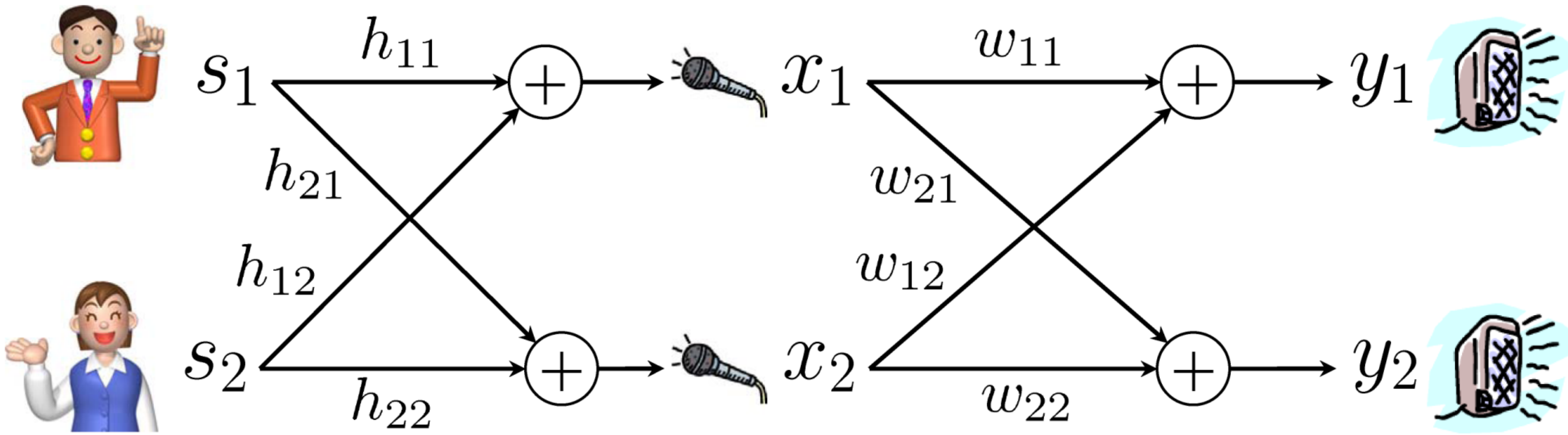
応用紹介：音源分離



a	b	a	b	c	a	b	c	d
x1	x2	x1	x2	x3	x1	x2	x3	x4
y1	y2	y1	y2	y3	y1	y2	y3	y4

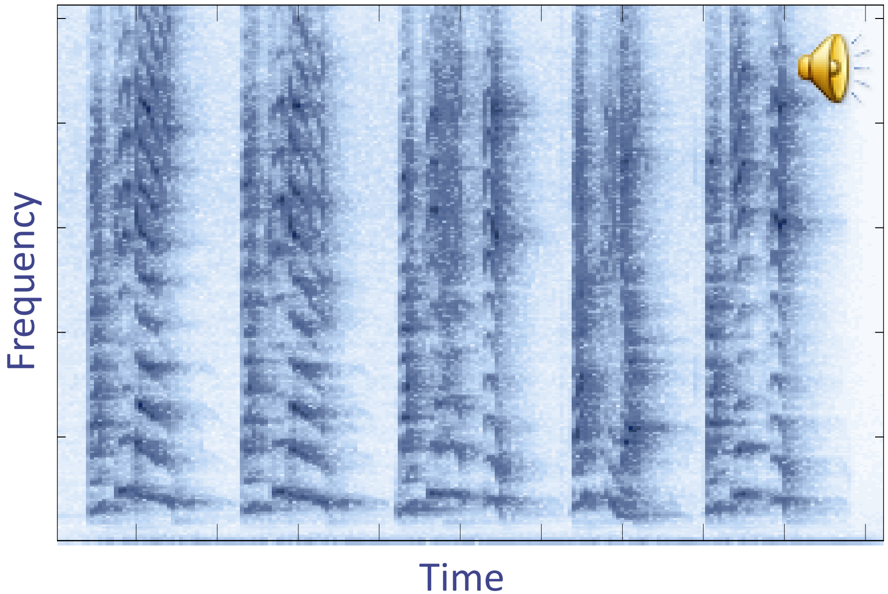
Cf) ブラインド音源分離（Blind Source Separation）

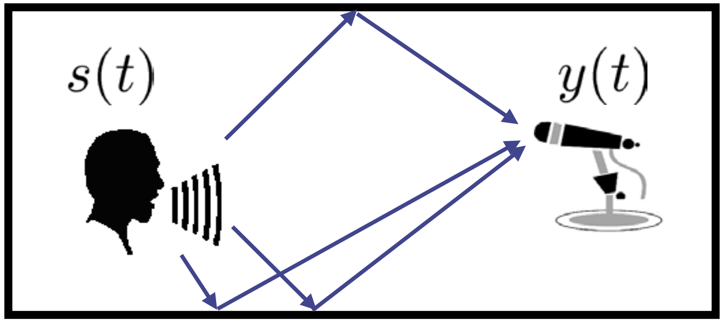
- 混ざり合った信号 x_1，x_2 から元の信号を取り出す
- どのように混ざったかに関する情報 $\boldsymbol{H}$ は利用できない



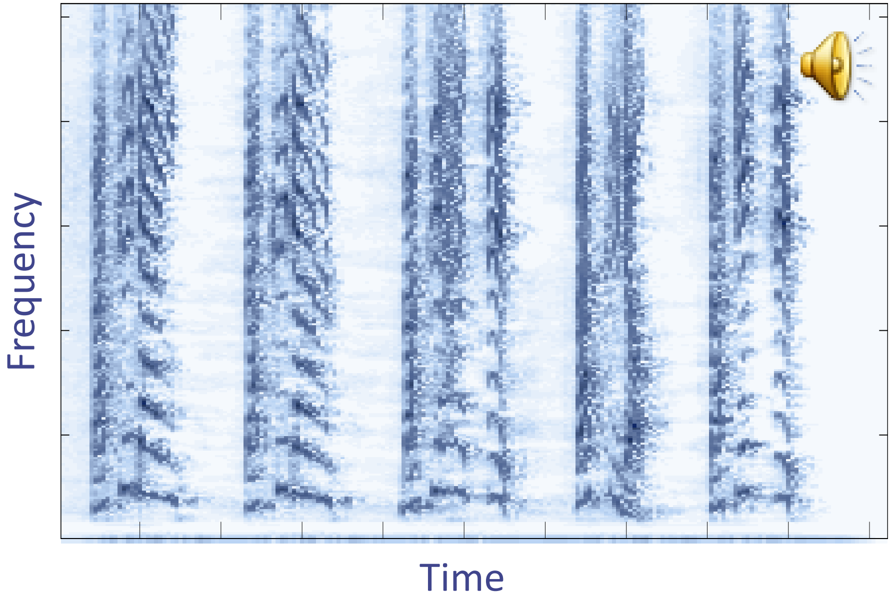
応用紹介：残響抑圧

■観測信号のスペクトログラム

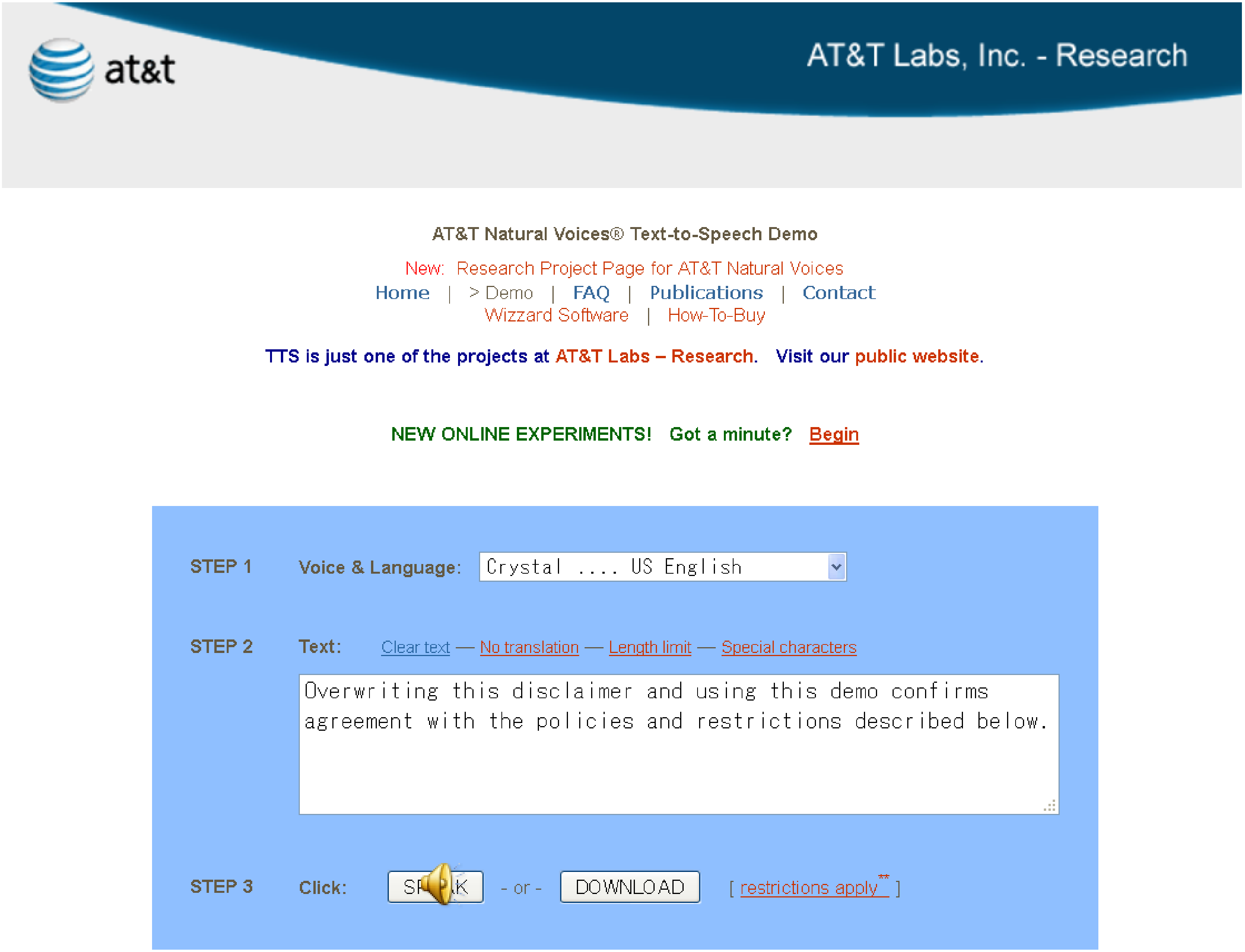




■残響除去信号のスペクトログラム



応用紹介：音声合成


at&t
AT&T Labs, Inc. - Research
AT&T Natural Voices® Text-to-Speech Demo
New: Research Project Page for AT&T Natural Voices
Home | > Demo | FAQ | Publications | Contact
Wizzard Software | How-To-Buy
TTS is just one of the projects at AT&T Labs – Research. Visit our public website.
NEW ONLINE EXPERIMENTS! Got a minute? Begin
STEP 1 Voice & Language: Crystal .... US English
STEP 2 Text: Clear text — No translation — Length limit — Special characters
Overwriting this disclaimer and using this demo confirms agreement with the policies and restrictions described below.
STEP 3 Click: SPEAK - or - DOWNLOAD [restrictions apply**]

確率則

- 同時確率（または結合確率）
  - $p(x,y)$: 事象xと事象yが同時に起こる確率
- 条件つき確率
  - $p(y|x)$: 事象xが起こった下で事象yが起こる確率

$$p(x,y) = p(y|x)p(x) = p(x|y)p(y)$$

- 周辺化
  - $p(y) = \int p(x,y)\mathrm{d}x$
- 独立性
  - xとyが独立 ⇔ $p(x,y) = p(x)p(y)$

ベイズの定理

- $$\underline{p(x|y)} = \frac{p(y|x)p(x)}{p(y)} = \frac{\underline{p(y|x)}\,\underline{p(x)}}{\int \underline{p(y|x)}\,\underline{p(x)}\mathrm{d}x}$$

x : 雨が降っている
y : 太郎が傘をもっている

↓

太郎が傘を持って現れた時外で雨が降っている確率 $\underline{p(X=x|Y=y)}$ が,
普段雨が降ると太郎が傘を持って出かける確率 $\underline{p(Y=y|X=x)}$ と,
雨が降る確率 $\underline{p(X=x)}$ を使って計算できる

確率モデル(尤度関数，事前確率)

- $$\underline{p(\theta|Y)} = \frac{\underline{p(Y|\theta)}\,\underline{p(\theta)}}{\int_\theta \underline{p(Y|\theta)}\,\underline{p(\theta)}}$$

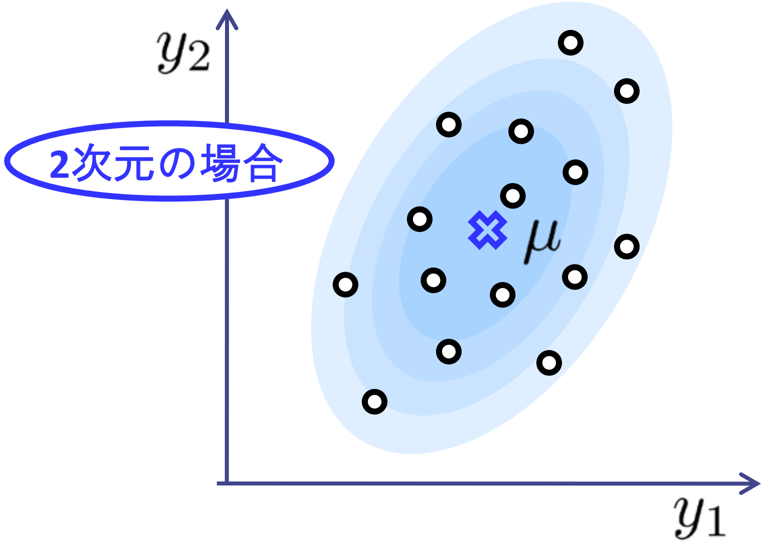
- 観測データを Y 、未知パラメータを θ とすると・・・
  - $\underline{p(Y|\theta)}$ のことを尤度関数
  - $\underline{p(\theta)}$ のことを事前確率
  - $\underline{p(\theta|Y)}$ のことを事後確率

という

確率モデル

- データ $\boldsymbol{y}$ の確率的な「生成源」 $p(\boldsymbol{y}|\boldsymbol{\theta})$

例1） 正規分布 (Normal distribution)

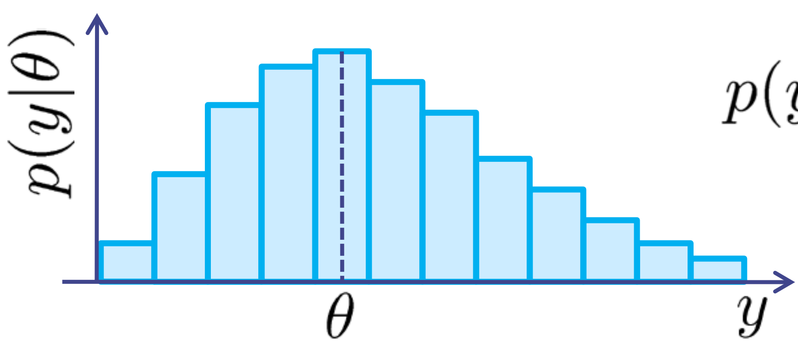


$$p(\boldsymbol{y}|\underbrace{\boldsymbol{\theta}}_{\{\boldsymbol{\mu}, \boldsymbol{\Sigma}\}})$$

$$= \frac{1}{\sqrt{(2\pi)^K |\boldsymbol{\Sigma}|}} e^{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})}$$

⇒以後 $\mathcal{N}(\boldsymbol{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ と表記

例2） Poisson分布



$$p(y|\theta) = \frac{\theta^y e^{-\theta}}{y!} \quad (y \in \mathbb{N})$$

⇒以後 $\mathrm{Pois}(y; \theta)$ と表記

正規分布に従う確率変数の諸性質

- $x \sim \mathcal{N}(x; \mu_x, \Sigma_x),\ y \sim \mathcal{N}(y; \mu_y, \Sigma_y) \cdots$
  ※「 $\sim$ 」は「従う」を意味する
  - x と y が独立なら $z = x + y \quad \Rightarrow \quad z \sim \mathcal{N}(\mu_x + \mu_y, \Sigma_x + \Sigma_y)$
  - x の線形変換は正規分布に従う
    $$z = Ax + b \quad \Rightarrow \quad z \sim \mathcal{N}(z; A\mu_x + b, A\Sigma_x A^\mathsf{T})$$
  - x と y の結合ベクトルは正規分布に従う
    $$\begin{bmatrix} x \\ y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_y \end{bmatrix}\right)$$

    このときの, y が与えられた下での x の条件つき期待値
    $$\mathbb{E}[x|y] = \mu_x + \Sigma_{xy}\Sigma_{yy}^{+}(y - \mu_y)$$

Poisson分布に従う確率変数の諸性質

- $x \sim \mathrm{Pois}(x;\lambda_x),\ y \sim \mathrm{Pois}(y;\lambda_y)$のとき・・・

  ※「 $\sim$ 」は「従う」を意味する

  - x と y が独立なら $z = x + y \quad \Rightarrow \quad z \sim \mathrm{Pois}(z;\lambda_x + \lambda_y)$
  - x と y が独立で $c = x + y$ のとき $x|c \sim \mathrm{Binom}(x;c,\lambda_x/(\lambda_x + \lambda_y))$

    $y|c \sim \mathrm{Binom}(y;c,\lambda_y/(\lambda_x + \lambda_y))$

パラメータ推定

データの確率的な生成プロセスの仮定（順問題）

⇔観測データの確率モデル化

データから生成プロセスのパラメータの推定（逆問題）

⇔最尤推定，最大事後確率推定，
最小平均二乗誤差推定，ベイズ推論

$p(y|\theta)$，$p(\theta)$ をモデル化

$$\theta \;\underset{\text{逆問題}}{\overset{\text{順問題}}{\rightleftarrows}}\; y$$

ベイズの定理

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)\mathrm{d}\theta}$$

ML推定量，MAP推定量，MMSE推定量

- データ y が与えられた下でのパラメータ θ の推定量
  - 最尤(Maximum Likelihood)推定量

$$\hat{\theta}_{\mathrm{ML}} = \operatorname*{argmax}_{\theta} p(Y|\theta)$$

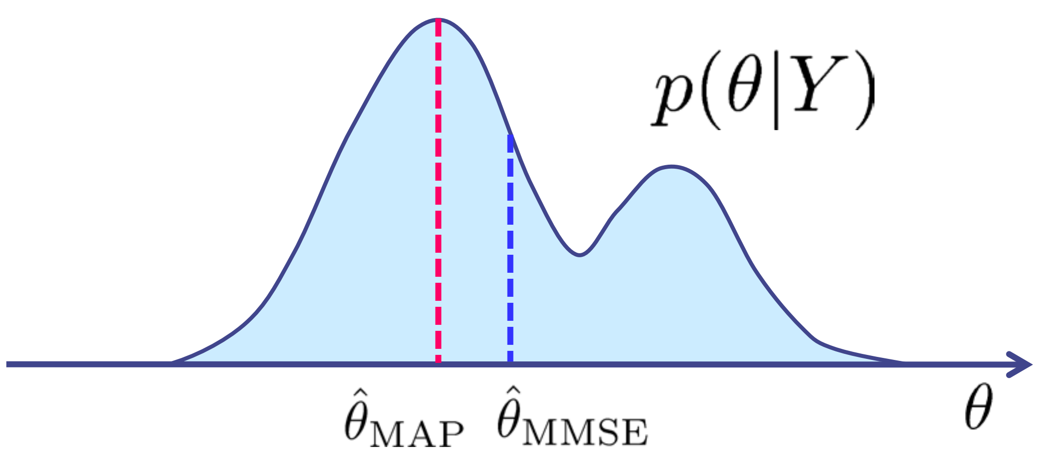
MAP推定で $p(\theta) \propto 1$ を仮定した場合に相当

  - 最大事後確率(Maximum A Posteriori)推定量

$$\hat{\theta}_{\mathrm{MAP}} = \operatorname*{argmax}_{\theta} p(\theta|y) = \operatorname*{argmax}_{\theta} p(y|\theta)p(\theta)$$

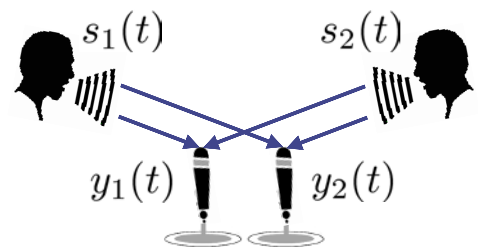
  - 最小平均二乗誤差(Minimum Mean Squared Error)推定量

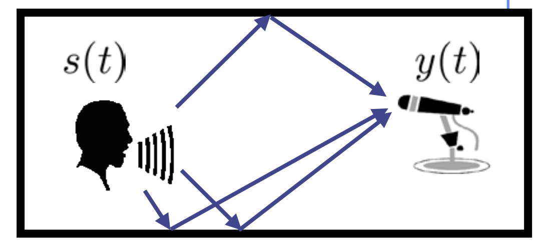
$$\hat{\theta}_{\mathrm{MMSE}} = \operatorname*{argmin}_{\hat{\theta}} \mathbb{E}\big[(\hat{\theta} - \theta)^2|Y\big] = \mathbb{E}\big[\theta|Y\big]$$

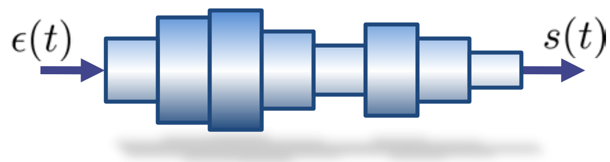


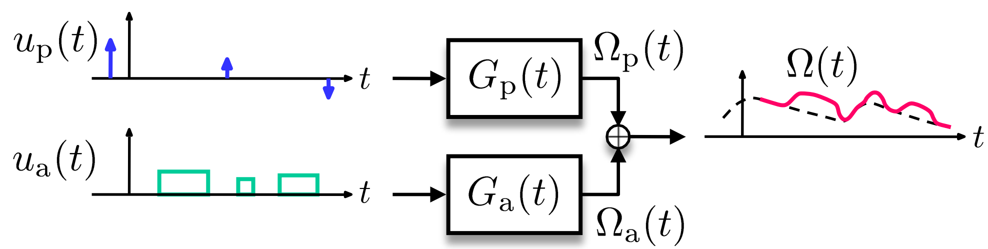
音声音響信号処理問題の多くは逆問題

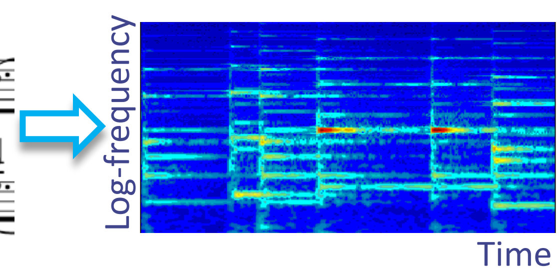
- 音響信号処理
  - ブラインド音源分離
  - 残響除去
  - ・・・
- 音声情報処理
  - 音素特徴抽出
  - 音声認識
  - イントネーション解析
- 音楽情報処理
  - 多重音解析
  - 自動採譜
  - ・・・











生成モデルアプローチ

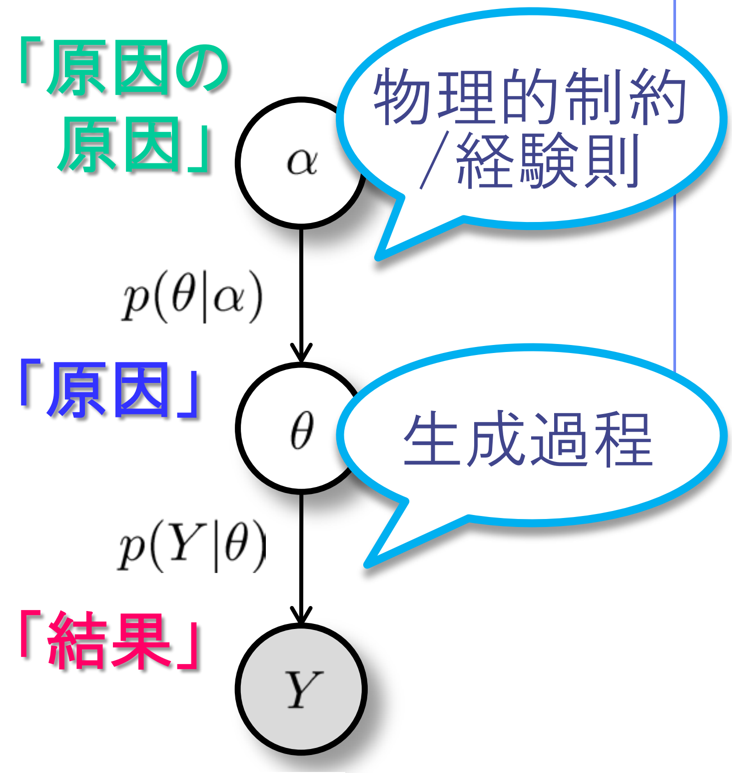
① 尤度関数の仮定

- 観測データ Y を生成する確率的なプロセス $p(Y|\theta)$ をモデル化

② 事前分布の仮定

- 生成モデルのパラメータθの生成プロセス $p(\theta)$ をモデル化

生成モデル（順問題）



③ 推論（逆問題）

- データ Y からθとαを推論
- 最尤推定量 $\hat{\theta} = \underset{\theta}{\mathrm{argmax}}\, p(Y|\theta)$，MAP推定量 $\hat{\theta} = \underset{\theta}{\mathrm{argmax}}\, p(\theta, \alpha|Y)$
MMSE推定量 $\hat{\theta} = \mathbb{E}[\theta, \alpha|Y]$，ベイズ事後分布 $p(\theta, \alpha|Y)$

Wienerフィルタの問題設定

- 問題：
  - $Y(\omega),\ X(\omega),\ N(\omega)$：観測信号，音声信号，雑音信号の離散Fourier変換（複素スペクトル）
  - 雑音重畳音声

$$Y(\omega) = X(\omega) + N(\omega)$$

  から音声に関係するパラメータ θ を推定したい

- 仮定：
  - 音声 $X(\omega)$ と雑音 $N(\omega)$ は無相関
  - 音声 $X(\omega)$ は平均0の複素正規分布 $\mathcal{N}_{\mathbb{C}}(0, \lambda_X(\omega))$ に従う
  - 雑音 $N(\omega)$ は平均0の複素正規分布 $\mathcal{N}_{\mathbb{C}}(0, \lambda_N(\omega))$ に従う
  - 雑音パワースペクトル密度 $\lambda_N(\omega)$ は既知
    （例えば無音声区間から推定済みという状況を想定）

$X(\omega)$ のMMSE推定量

- 問題設定：
  - $Y(\omega) = X(\omega) + N(\omega)$
  - $N(\omega) \sim \mathcal{N}_{\mathbb{C}}(N(\omega); 0, \lambda_N(\omega)), \quad X(\omega) \sim \mathcal{N}_{\mathbb{C}}(X(\omega); 0, \lambda_X(\omega))$
- 求めたいのは $\hat{X}_{\mathrm{MMSE}}(\omega) = \mathbb{E}[X(\omega)|Y(\omega)]$

多変量Gauss分布の性質

$$p(y, x) = \mathcal{N}\left(\begin{bmatrix} y \\ x \end{bmatrix}; \begin{bmatrix} \bar{y} \\ \bar{x} \end{bmatrix}, \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix}\right) \longrightarrow \mathbb{E}[x|y] = \bar{x} + \Sigma_{xy}\Sigma_{yy}^{+}(y - \bar{y})$$

$$p(Y(\omega), X(\omega)) = \mathcal{N}_{\mathbb{C}}\left(\begin{bmatrix} Y(\omega) \\ X(\omega) \end{bmatrix}; \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \lambda_X(\omega) + \lambda_N(\omega) & \lambda_X(\omega) \\ \lambda_X(\omega) & \lambda_X(\omega) \end{bmatrix}\right)$$

$$\longrightarrow \mathbb{E}[X(\omega)|Y(\omega)] = \frac{\lambda_X(\omega)}{\lambda_X(\omega) + \lambda_N(\omega)} Y(\omega)$$ （Wienerフィルタ）

離散時間Kalmanフィルタの問題設定

システムモデル： $x_k = \Phi_{k,k-1} x_{k-1} + \underbrace{B_k v_k}_{\text{駆動雑音}}$

測定モデル： $y_k = C_k x_k + \underbrace{w_k}_{\text{観測雑音}}$

k ： 時刻インデックス

- 仮定
  - v_k， w_k は互いに独立な正規白色雑音

$$\mathbb{E}[v_k] = 0 \qquad \mathbb{E}[v_k v_n^\top] = V_k \delta_{kn}$$

$$\mathbb{E}[w_k] = 0 \qquad \mathbb{E}[w_k w_n^\top] = W_k \delta_{kn}$$

$$\mathbb{E}[v_k w_n^\top] = 0$$

  - パラメータ：$\Phi_{k,k-1}, B_k, C_k$ と、雑音共分散 V_k, W_k は既知

離散時間Kalmanフィルタの構成

状態推定値

$$\hat{x}_{k-1|k-1} \quad \Rightarrow \quad \hat{x}_{k|k-1} \quad \Rightarrow \quad \hat{x}_{k|k}$$

時刻k-1までの観測値を用いた時刻k-1の状態推定値

時間更新

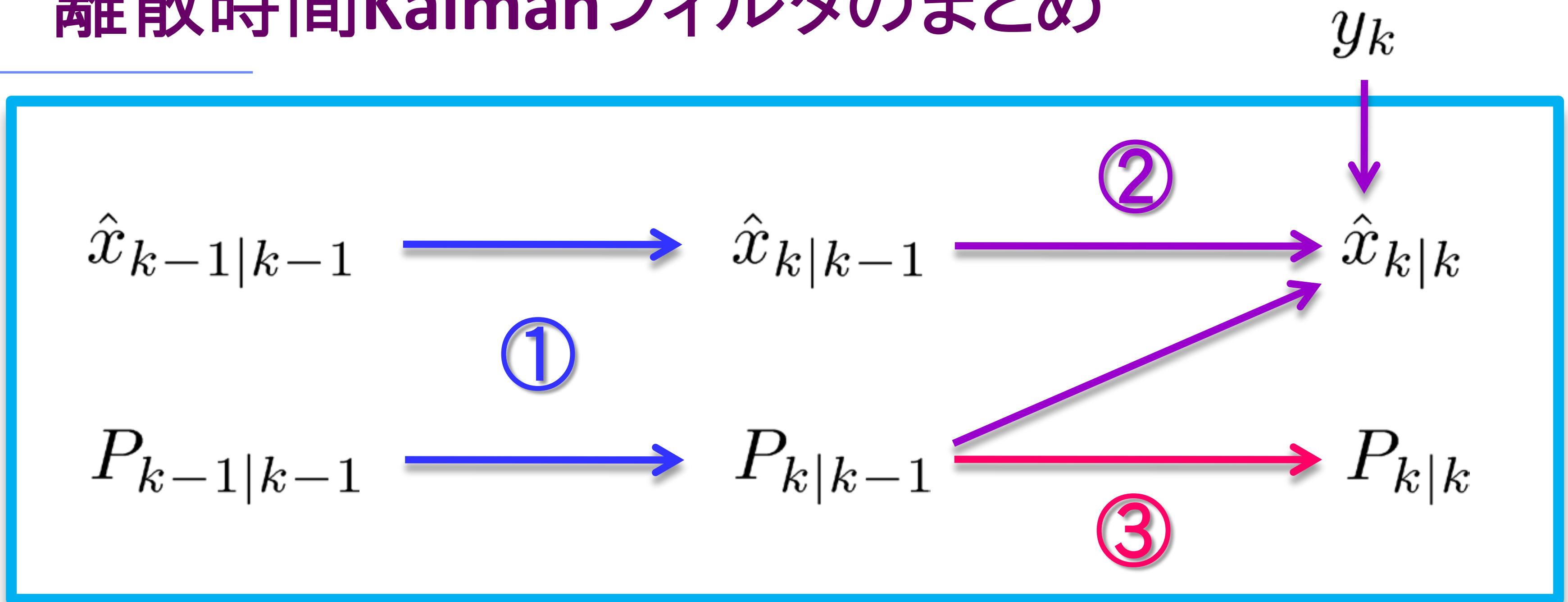
時刻k-1までの観測値を用いた時刻kの状態推定値

計測更新

時刻kまでの観測値を用いた時刻kの状態推定値

観測値 y_k

離散時間Kalmanフィルタのまとめ



①
$$\hat{x}_{k|k-1} = \Phi_{k,k-1}\hat{x}_{k-1|k-1}$$
$$P_{k|k-1} = \Phi_{k,k-1}P_{k-1|k-1}\Phi_{k,k-1}^{\top} + B_k V_k B_k^{\top}$$

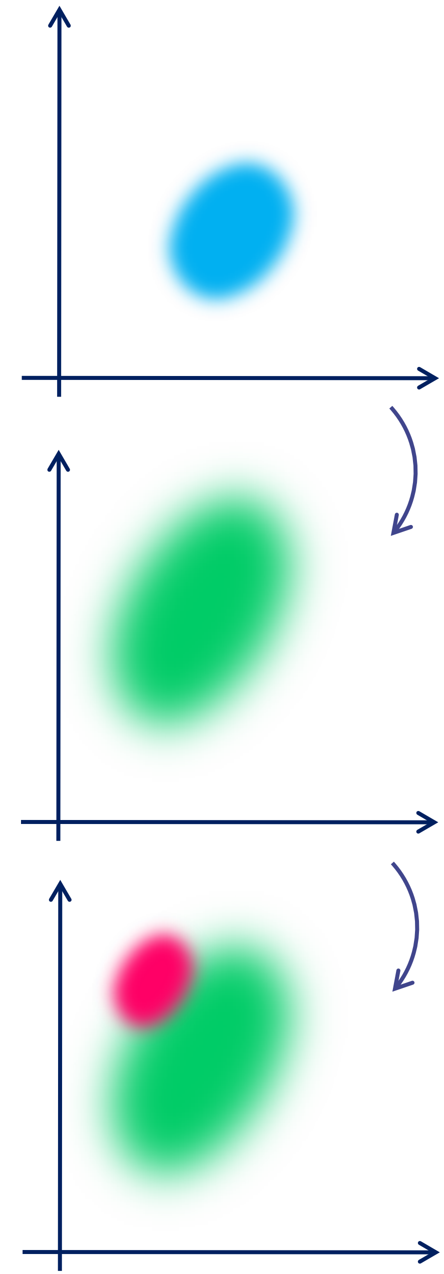
②
$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + P_{k|k-1}C_k^{\top}(C_k P_{k|k-1} C_k^{\top} + W_k)^{-1}(y_k - C_k\hat{x}_{k|k-1})$$

③
$$P_{k|k} = (I - P_{k|k-1}C_k^{\top}(C_k P_{k|k-1} C_k^{\top} + W_k)^{-1}C_k)P_{k|k-1}$$

逐次更新アルゴリズムで計算される確率分布

- $p(x_{k-1}|y_{1:k-1})$
  - 時刻t_1からt_{k-1}までの観測信号が与えられたもとでの時刻t_{k-1}における状態推定値の事後確率分布

  $\hat{x}_{k-1|k-1}, P_{k-1|k-1}$:この分布の平均と共分散行列に相当
- $p(x_k|y_{1:k-1})$
  - 時刻t_{k-1}における上記事後分布を手がかりに推測される、時刻t_kにおける状態推定値の事前確率分布
- $p(x_k|y_{1:k}) = p(x_k|y_k, y_{1:k-1})$
  - 上記事前分布と時刻t_kにおける観測信号をもとに得られる、時刻t_kにおける状態推定値の事前確率分布



離散時間Kalmanフィルタのベイズ的解釈

$$\boxed{p(x_{k-1}|y_{1:k-1})} = \mathcal{N}(\hat{x}_{k-1|k-1}, P_{k-1|k-1})$$

システムモデルより:

$$\underline{p(x_k|x_{k-1})} = \mathcal{N}(\Phi_{k,k-1}x_{k-1}, B_k V_k B_k^{\mathsf{T}})$$

計測モデルより:

$$\underline{p(y_k|x_k)} = \mathcal{N}(C_k x_k, W_k)$$

時間更新 $$\underline{p(x_k|y_{1:k-1})} = \int \underline{p(x_k|x_{k-1})}\boxed{p(x_{k-1}|y_{1:k-1})}\mathrm{d}x_{k-1}$$

計測更新 $$\boxed{p(x_k|y_{1:k})} = \frac{\underline{p(y_k|x_k)}\,\underline{p(x_k|y_{1:k-1})}}{p(y_k|y_{1:k-1})}$$

$$= \frac{\underline{p(y_k|x_k)}\,\underline{p(x_k|y_{1:k-1})}}{\int \underline{p(y_k|x_k)}\,\underline{p(x_k|y_{1:k-1})}\mathrm{d}x_k}$$

$$\boxed{p(x_k|y_{1:k})} = \mathcal{N}(\hat{x}_{k|k}, P_{k|k})$$