

信号処理論特論

第1回 (4/14)

情報理工学系研究科システム情報学専攻
猿渡 洋

hiroshi_saruwatari@ipc.i.u-tokyo.ac.jp

講義予定

- 04/14: 第1回 統計的音声音響信号処理概論
- 04/21: 第2回 非負値行列因子分解
- 04/28: 第3回 ブラインド音源分離その1
- 05/12: 第4回 ブラインド音源分離その2
- 05/19: 第5回 エンハンスメント・高次統計量解析とその応用
- 05/26: 第6回 【レポート課題1】
- 06/02: 第7回 音声合成・変換その1
- 06/09: 第8回 音声合成・変換その2
- 06/16: 第9回 音場の計測と制御
- 06/23: 第10回 音楽情報処理
- 06/30: 第11回 【レポート課題2】

講義資料と成績評価

■講義資料

- <http://www.sp.ipc.i.u-tokyo.ac.jp/~saruwatari/>

（システム情報第一研究室からたどれるようにしておきます）

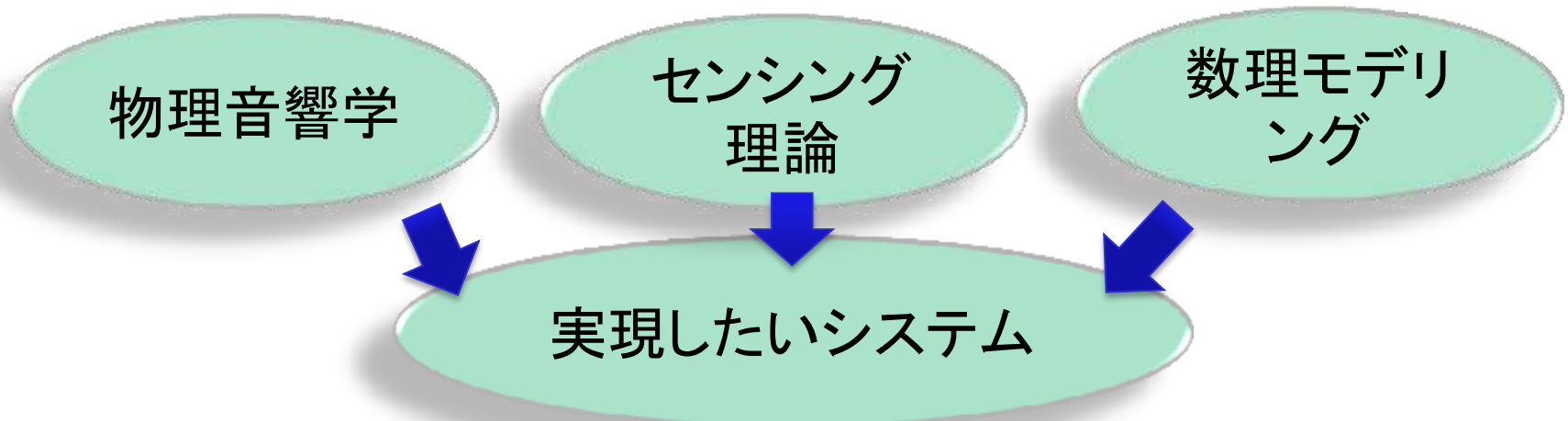
■成績評価

- レポート点（2回の提出が必須）

統計的音声音響信号处理概論

なぜ音メディアに焦点をあてるのか？

- 音メディアに関する信号処理研究の魅力とは？
 - 自然界の音が持つ無限の**多様性** (cf. 無線通信信号)
 - 研究のアプローチに**多面性**あり (決定論的？統計的？)
 - 最後は聴かせてなんぼの評価 ⇒ **芸術性**も併せ持つ
- 「**物理世界 (波動) と情報世界 (抽象) をまたぐ学問**」であり、かつそれを「**統一的に取り扱うシステム工学**」である。
- 対象の多様性ゆえに「**なんでもあり**」の分野でもある。



なぜ音メディアに焦点をあてるのか？

本講義で特に扱うトピックス

波動方程式
室内音響
伝達関数
音生成過程 etc.

物理音響学

離散サンプリング
フーリエ解析
球面調波解析
圧縮センシング etc.

センシング
理論

統計モデリング
最尤・ベイズ推定
機械学習
スパース最適化

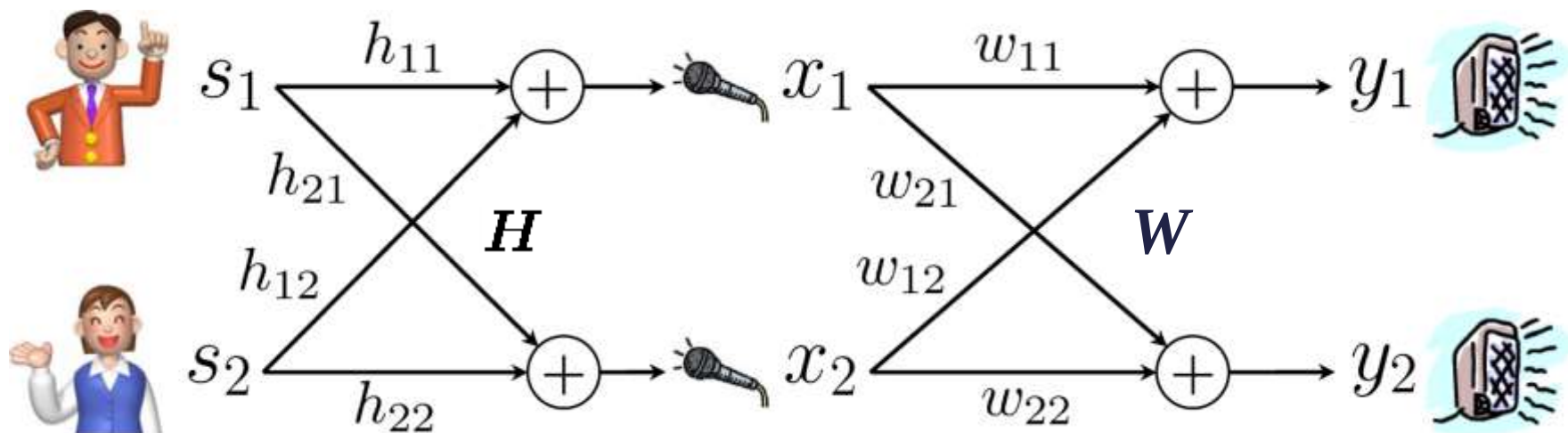
数理モデリ
ング

実現したいシステム



応用紹介: ブラインド音源分離 (BlindSourceSeparation)

- 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空間情報 H は利用できない



実は上記は**2つのことを同時に推定**している

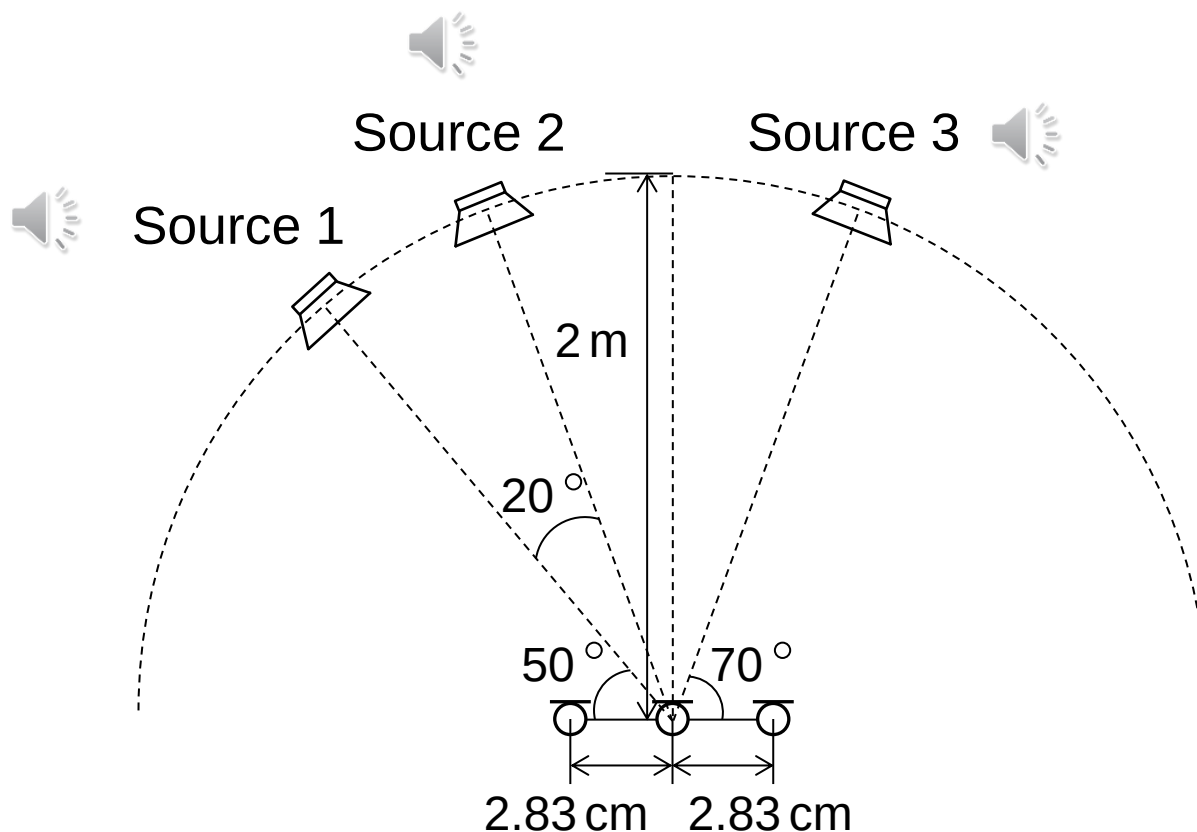
- [空間] 統計的に独立な音源の分類問題 (分離行列 W の推定)
- [信号] 各音源が属する確率分布 $p(y)$ の推定問題

上記を閉形式で解く方法は存在せず凸問題でもない⇒**大変困難!**

応用紹介：低ランク近似音源モデルに基づくBSS

[Kitamura, et al. 2015]

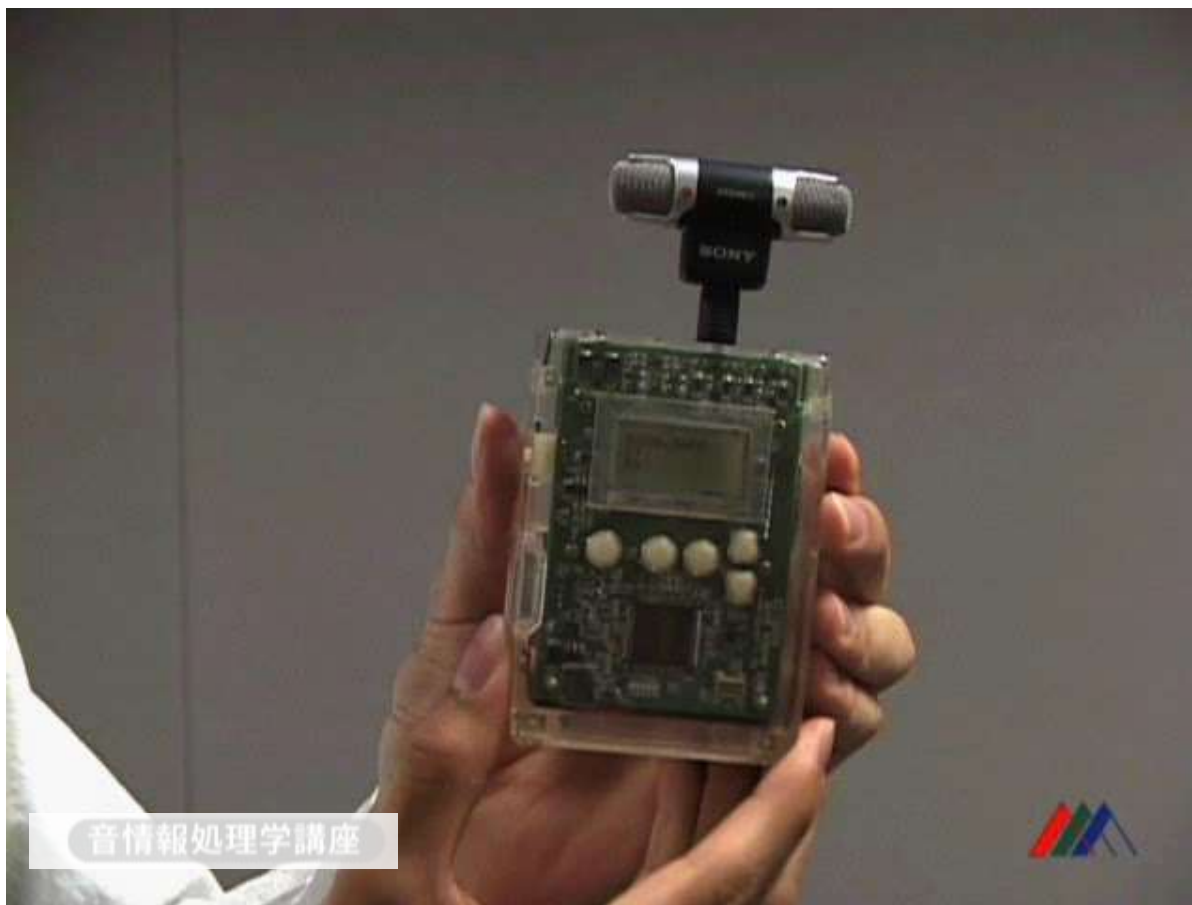
- ドラム、ストリング、音声からなる複合音の分離



応用紹介:「聖徳太子マイク」 リアルタイムBSS

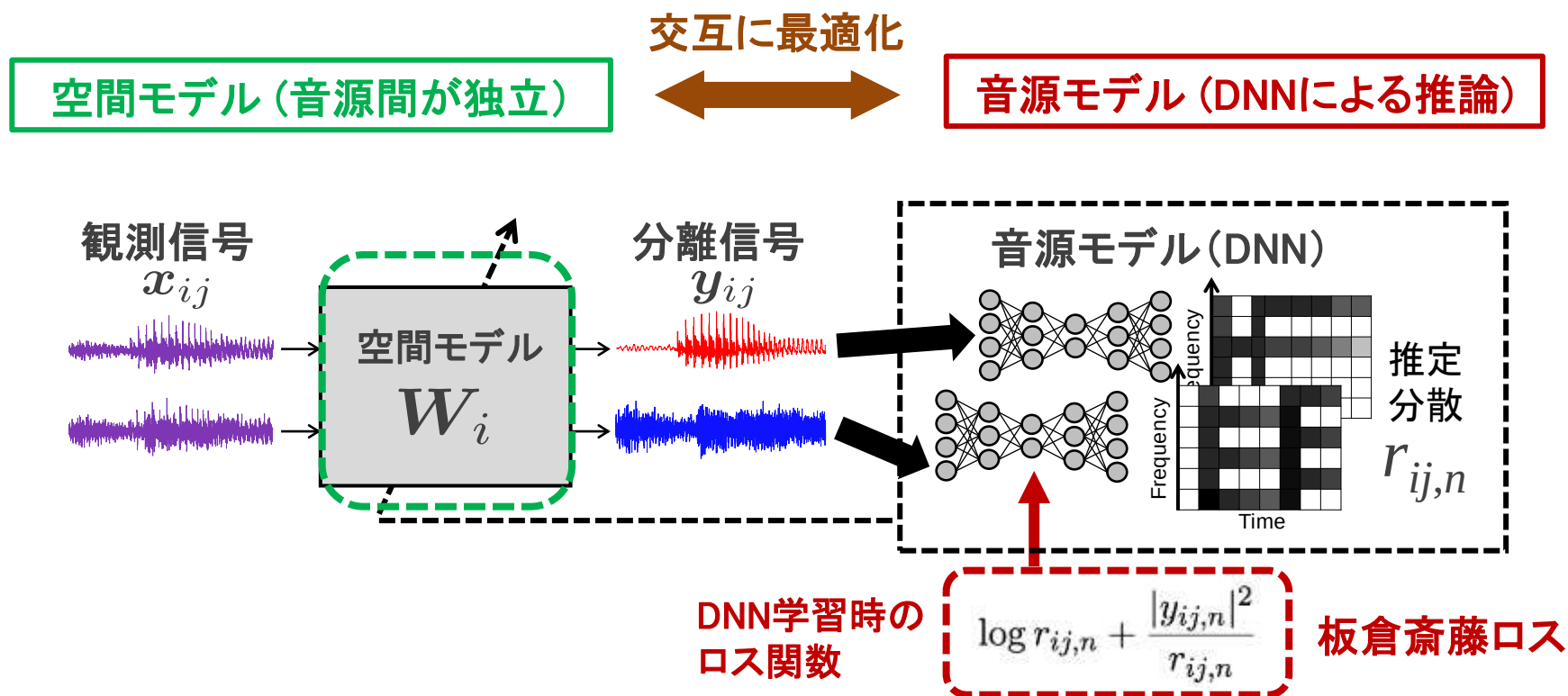
[Saruwatari, et al. 2009]

- 4chマイクとDSP(TI社製C67)から構成される独立成分分析BSS専用モジュールを2005年に開発
- 2009年に世界初の商用化(警察備品に採用)



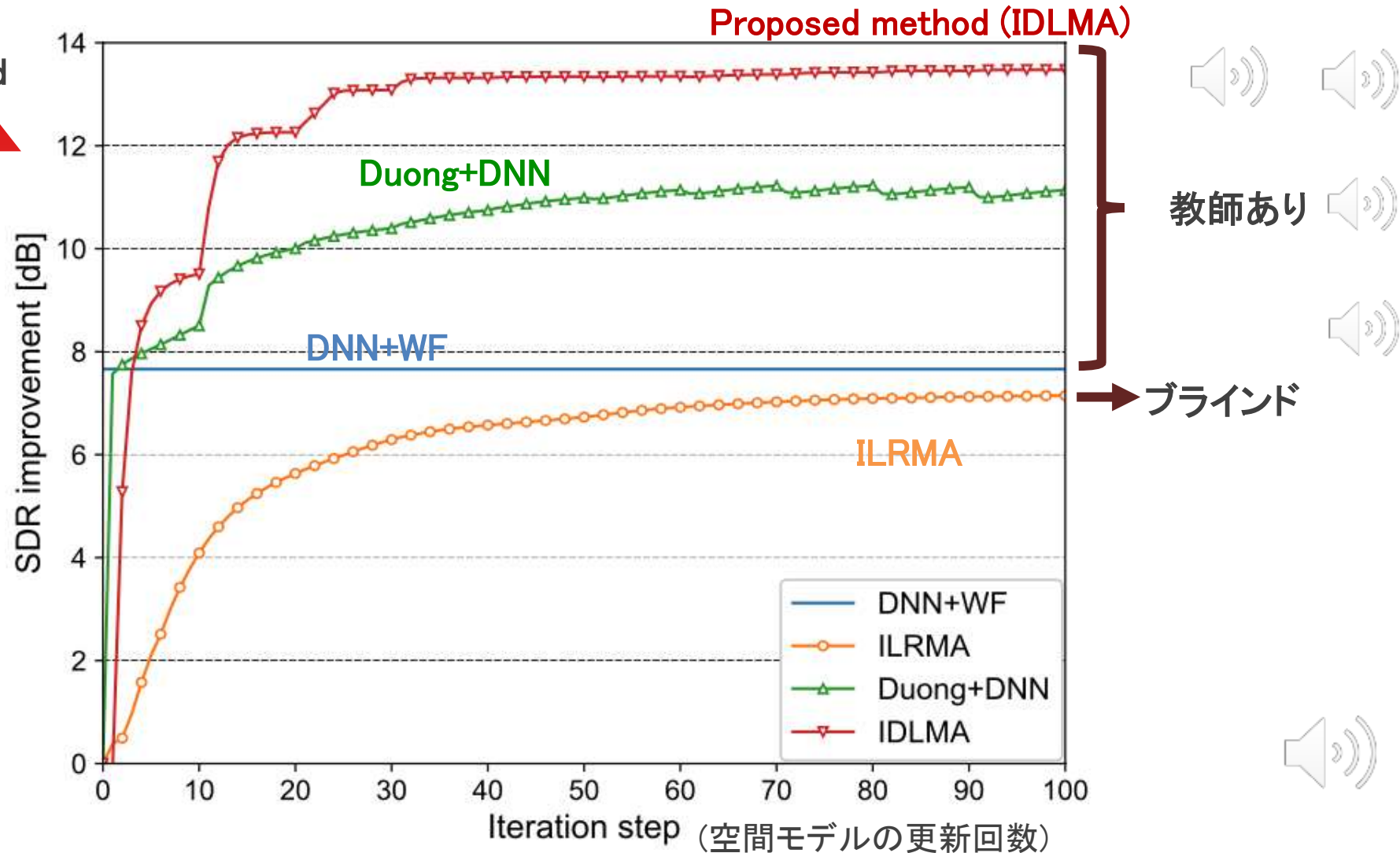
DNN(深層学習)による多チャネル音源分離

- 独立深層学習行列分析 (Independent Deeply Learned Matrix Analysis: IDLMA) [角野他 2018] 世界初の尤度最大化に基づくDNN空間音源分離



- **空間モデル**: 各音源が統計的に独立となる分離行列を推定
- **音源モデル**: 分散 $r_{ij,n}$ を推定するDNNを各音源ごとに構成

DNN(深層学習)による多チャネル音源分離

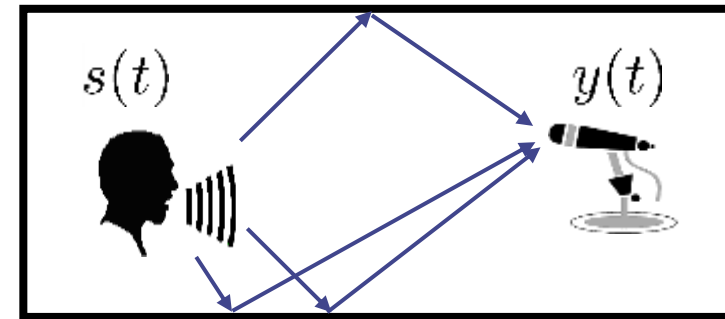
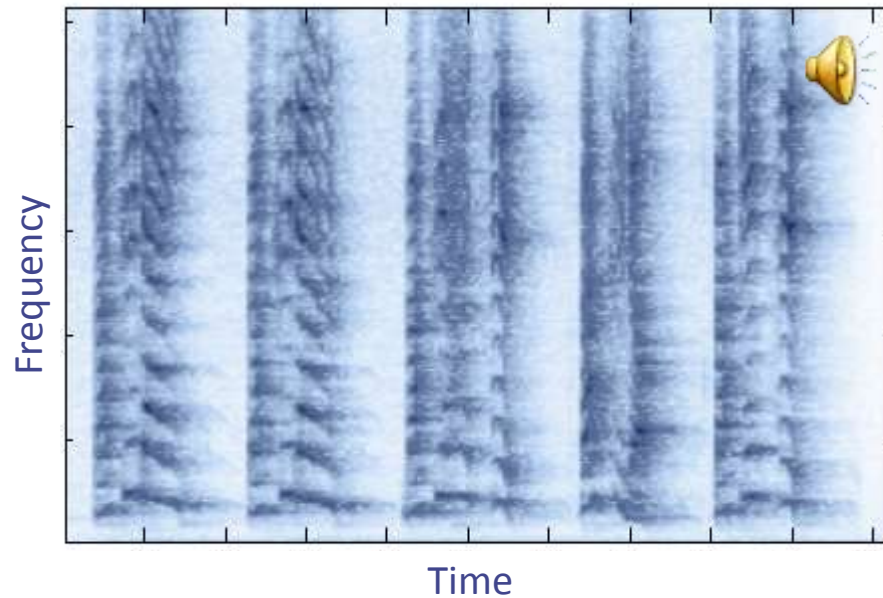


■ 10回に1回 DNNで分散行列を更新

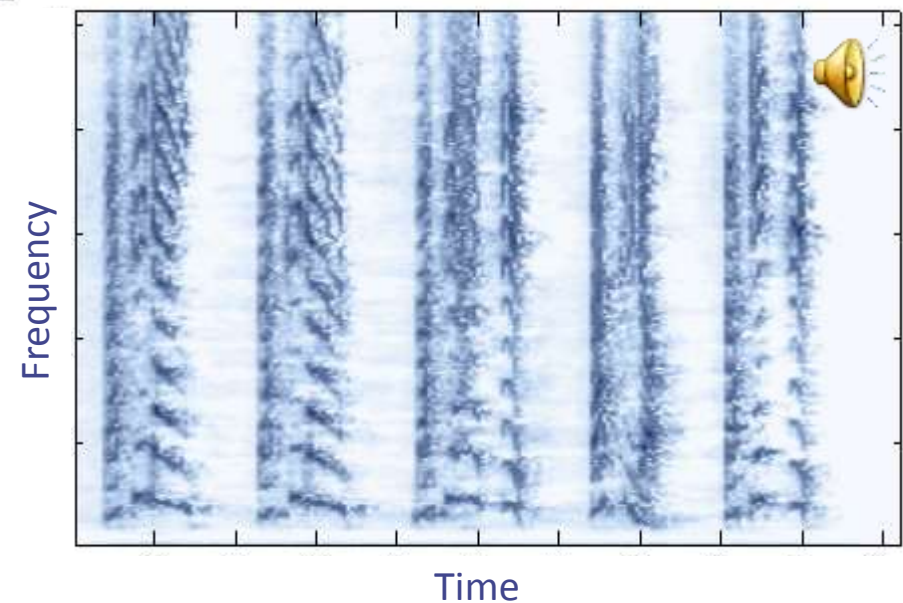
応用紹介: 残響抑圧

[Kameoka, et al.]

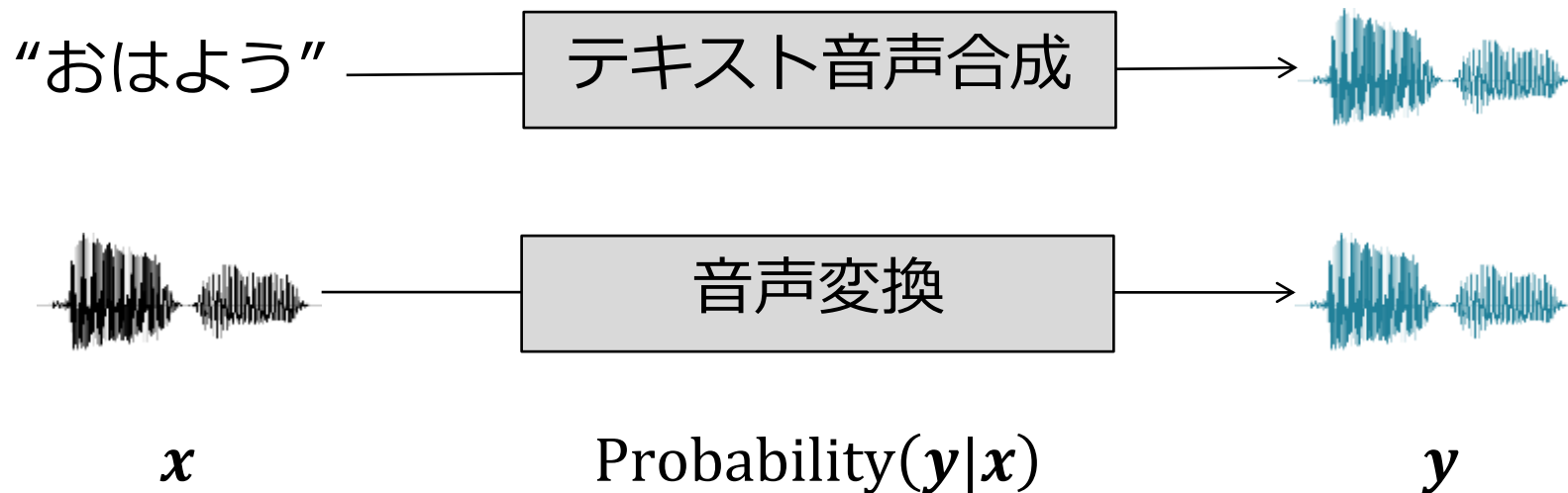
■ 観測信号のスペクトログラム



■ 残響除去信号のスペクトログラム



応用紹介: 音声合成



入力 x と出力 y の関係をどう記述するか？ → 逆問題

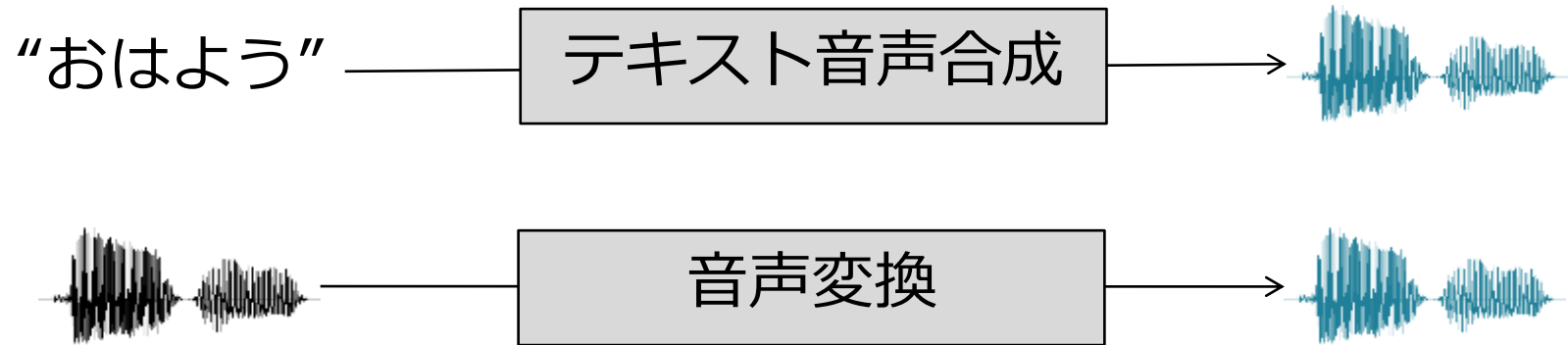
声のゆらぎをどう扱うか？

人間は二度と同じ声を出せない → 声色、音高、音の長さのゆらぎ

人間らしい声とは何か？

「人間らしい声」の定式化・統計モデル化

応用紹介: 音声合成



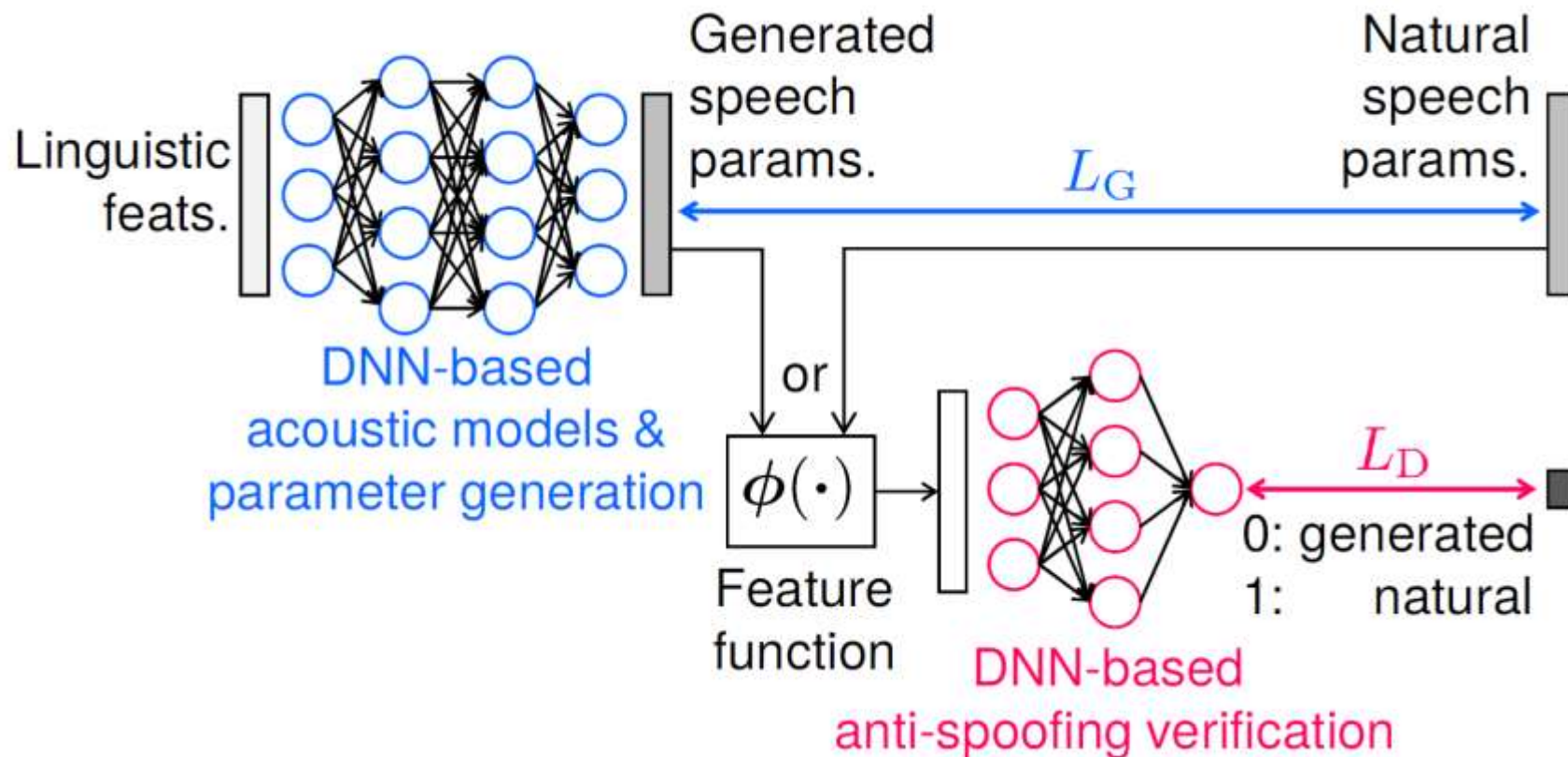
東京大学は世界一！
昼飯はとんこつラーメンに限る！





世界初のGAN（敵対学習）-DNN音声合成

[Saito, et al.; ICASSP2017 SLP Student Grant賞・TAF技術学生賞他]



人間の声に似せ
ようと努力



ウソ(合成音)に
騙されまいと攻防





世界初のGAN（敵対学習）-DNN音声合成

[Saito, et al.; ICASSP2017 SLP Student Grant賞・TAF技術学生賞他]



IEEE/ACM Transactions ASLP の表紙を飾りました

IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING

A PUBLICATION OF THE IEEE SIGNAL PROCESSING SOCIETY



Association for
Computing Machinery

www.signalprocessingsociety.org

Indexed in PubMed® and MEDLINE®, products of the United States National Library of Medicine



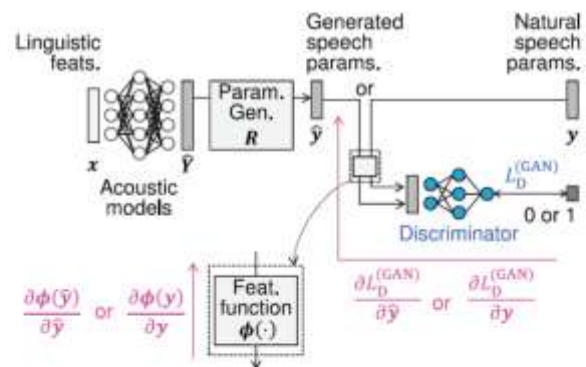
JANUARY 2018
FEBRUARY 2018

VOLUME 26
VOLUME 26

NUMBER 1
NUMBER 2

ITASFA

(ISSN 2329-9290)



Loss function and gradients for updating the discriminator. Param. Gen. indicates the speech parameter generation. For more, see "Statistical Parametric Speech Synthesis Incorporating Generative Adversarial Networks" by Y. Saito et al., p. 84.

For the January 2018 issue, see p. 1 for Table of Contents

For the February 2018 issue, see p. 210 for Table of Contents



ロボット対話 デモンストレーション

**Demonstration for
spoken-oriented hands-free
interactive robot.**

応用紹介: 統計的時系列推定における音質の差

[Saruwatari, et al. 2012]

白色ノイズの場合

観測音



最尤推定



ベイズ推定



ミュージカル
ノイズフリー



人ごみノイズの場合

観測音



最尤推定



ベイズ推定



ミュージカル
ノイズフリー



どの推定方式が音として「自然」か？

良い統計的性質を持つ推定法 $\neq$ 音響的に優れた技術

応用紹介：音バーチャルリアリティ

[Koyama, et al. 2014]

物理的なスピーカ列はここ



統計的信号処理の基礎

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

- $p(y) = \int p(x, y)dx$

■ 独立性

- x と y が独立 $\Leftrightarrow p(x, y) = p(x)p(y)$

ベイズの定理

$$\blacksquare \quad \underline{p(x|y)} = \frac{p(y|x)p(x)}{p(y)} = \frac{\underline{p(y|x)}\underline{p(x)}}{\int \underline{p(y|x)}\underline{p(x)}dx}$$

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)

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

■ 観測データを Y 、未知パラメータを θ とすると...

■ $p(Y|\theta)$ のことを尤度関数

■ $p(\theta)$ のことを事前確率

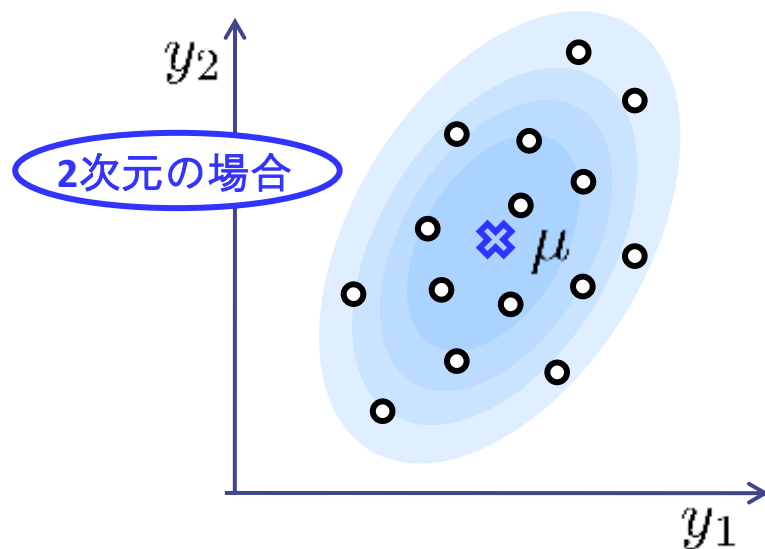
■ $p(\theta|Y)$ のことを事後確率

という

確率モデル

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

例1) 正規分布 (Normal distribution)

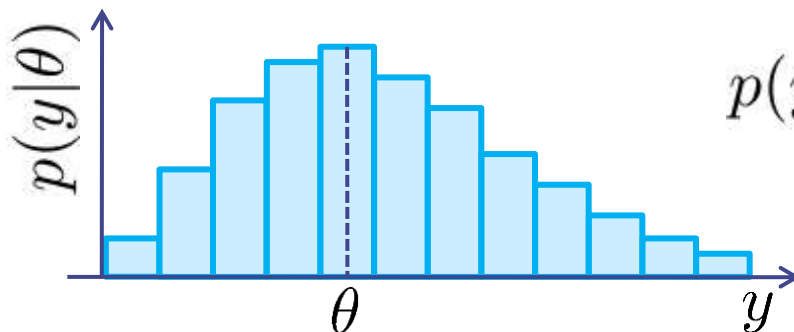


$\{\boldsymbol{\mu}, \boldsymbol{\Sigma}\}$ 平均、分散

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

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

例2) Poisson分布



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

⇒ 以後 $\text{Pois}(y; \theta)$ 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 $z = x + y \Rightarrow z \sim \mathcal{N}(\mu_x + \mu_y, \Sigma_x + \Sigma_y)$

■ x の線形変換は正規分布に従う

$$z = Ax + b \Rightarrow z \sim \mathcal{N}(z; A\mu_x + b, A\Sigma_x A^T)$$

■ 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

■ $x \sim \text{Pois}(x; \lambda_x)$, $y \sim \text{Pois}(y; \lambda_y)$ のとき...

※「 $\sim$ 」は「従う」を意味する

■ x と y が独立なら $z = x + y \Rightarrow z \sim \text{Pois}(z; \lambda_x + \lambda_y)$

■ x と y が独立で $c = x + y$ のとき $x|c \sim \text{Binom}(x; c, \lambda_x/(\lambda_x + \lambda_y))$

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

二項分布

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

順問題

θ $\longrightarrow$ y

逆問題

ベイズの定理

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

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

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

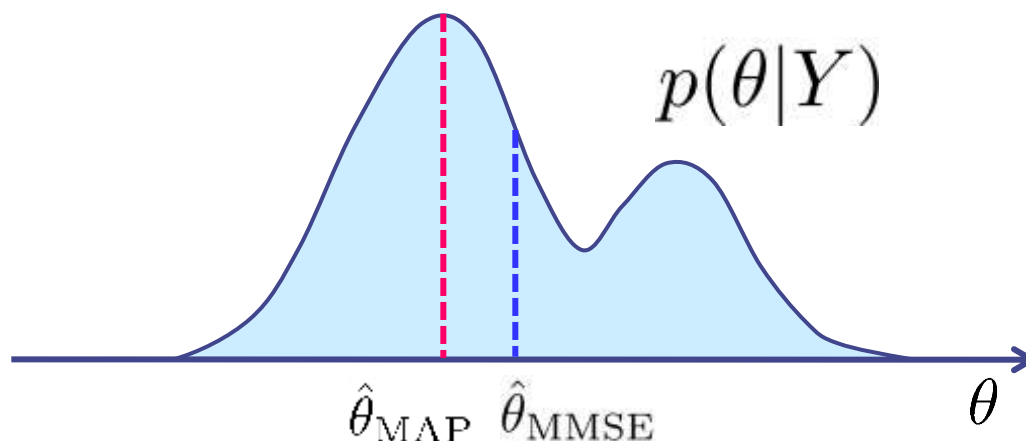
MAP推定で $p(\theta) \propto 1$
を仮定した場合に相当

■ 最大事後確率(Maximum A Posteriori)推定量

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

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

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

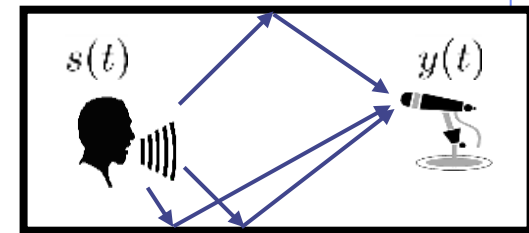
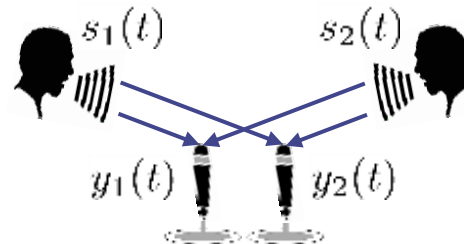


音声音響信号処理問題の多くは逆問題

■音響信号処理

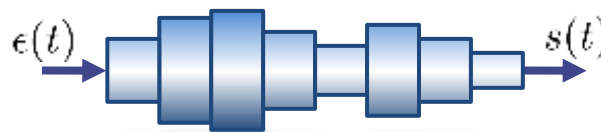
- ブラインド音源分離
- 残響除去

...



■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



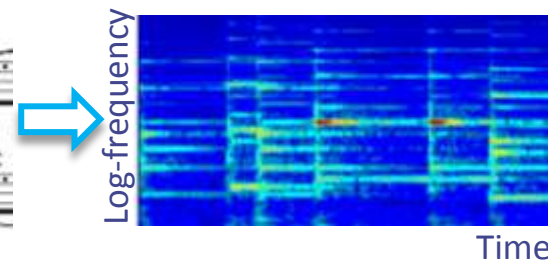
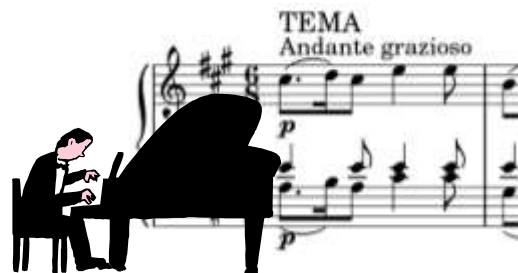
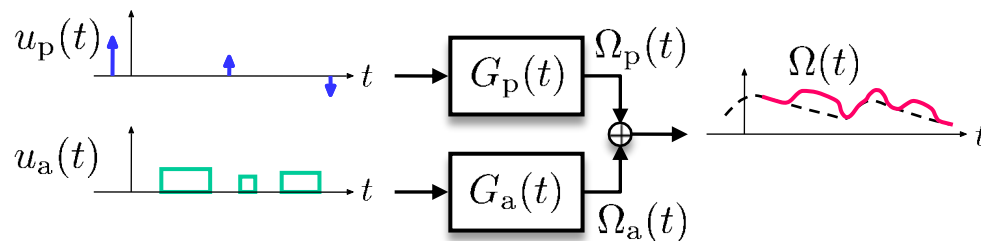
今日は晴天なり



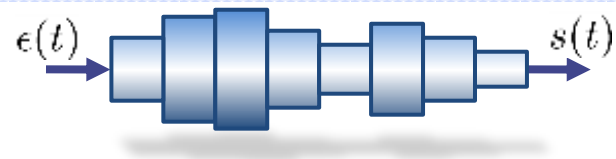
■音楽情報処理

- 多重音解析
- 自動採譜

...

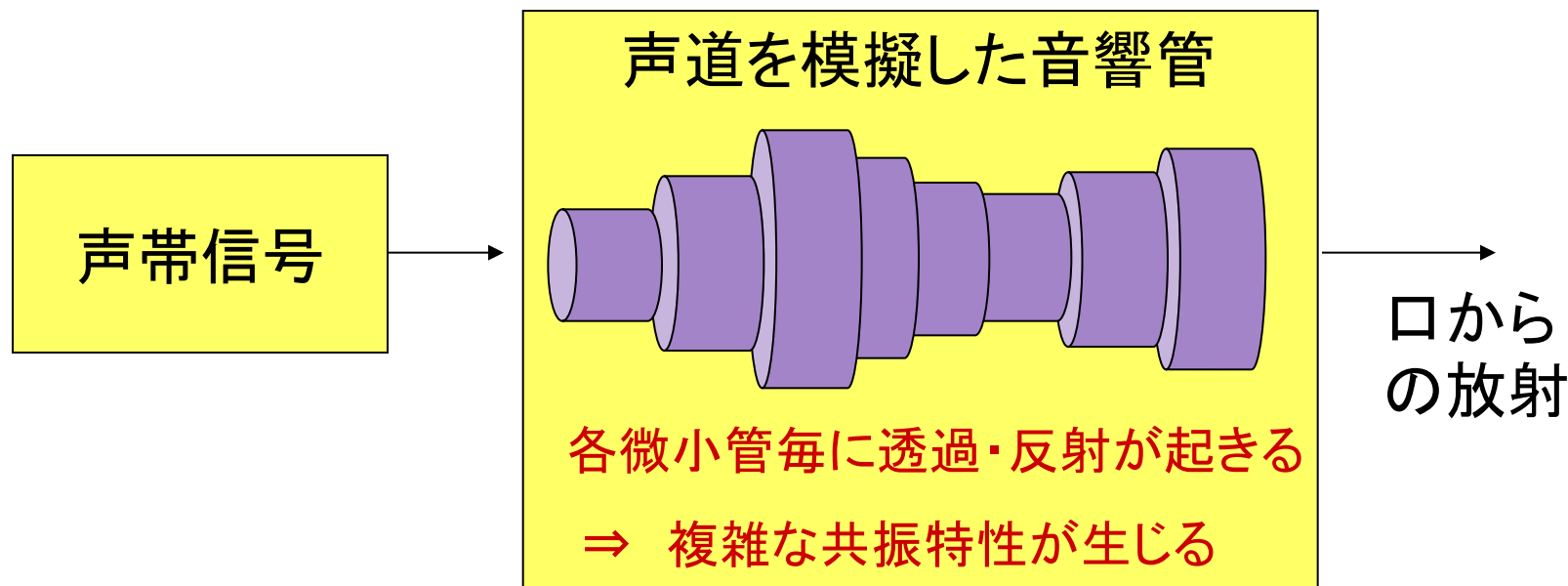


携帯電話における逆問題



■ 人間の音声生成モデル

- 声帯での基本振動を声道で音色付ける。
- 声道
 - 位置によって太さの異なる音響管の連続と見なせる。
 - 音響管における共振現象 ⇒ 自己回帰 (AR) 過程



この生成モデルパラメータを求めて伝送する⇒大幅な情報圧縮

音声のスペクトル構造

補足

■ 短時間スペクトル

- 音声は、短時間区間ごとの電力スペクトル密度（周波数領域におけるパワー特性）で測ることが多い。

■ 音声スペクトル構造の2要素

- 周波数とともにゆるやかに変化する成分[スペクトル包絡]
⇒発声器官の共振・反共振特性を表す
(つまり人間の喉・口の形をあらわす特徴量)
- 細かく周期的(有声音;母音などの場合)または非周期的(無声音の場合)に変化する成分 [スペクトル微細構造]
⇒音源の周期性
(つまり声帯の基本周期・声の高低を表す特徴量)

音声信号のスペクトルはこれら2つの要素の積で表される

音声のスペクトル構造

補足

人間頭部の断面図

②

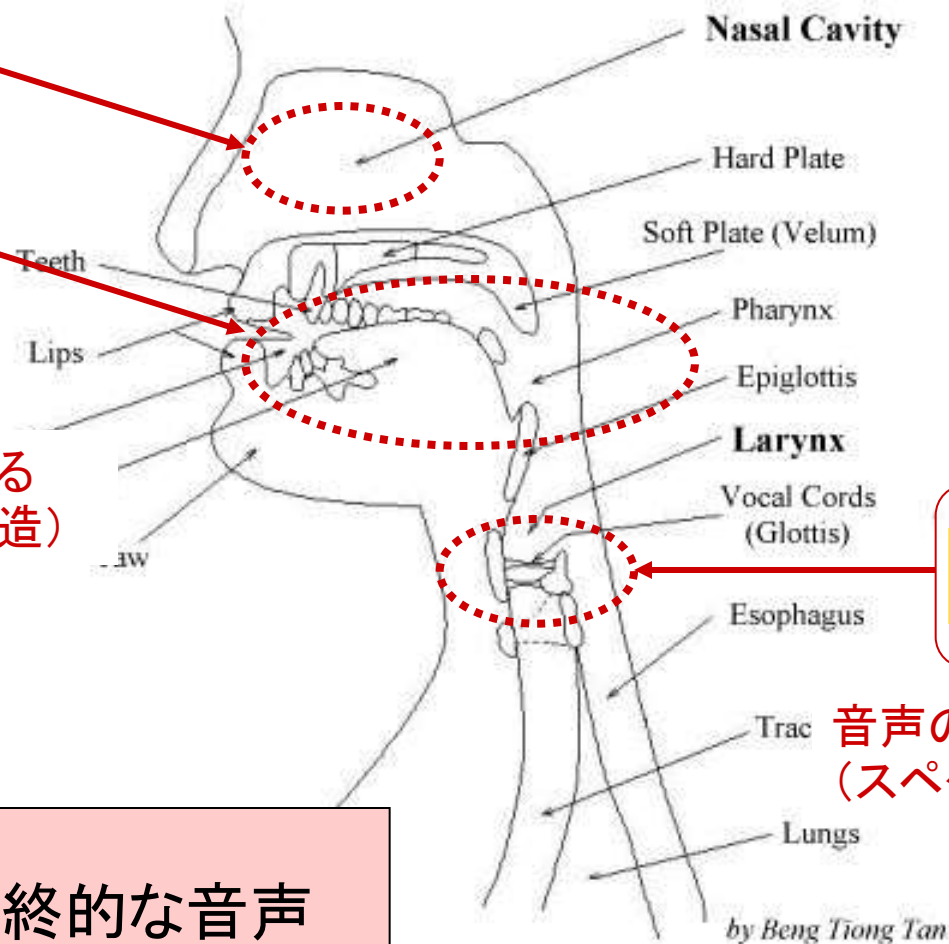
鼻腔

声道

の共振

音声の音色を与える
(スペクトル包絡構造)

① × ② = 最終的な音声



①

声帯の振動

音声の基本周期を与える
(スペクトル微細構造)

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

③ 推論(逆問題)

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

「原因の
原因」

物理的制約
/ 経験則

$p(\theta|\alpha)$

「原因」

生成過程

$p(Y|\theta)$

「結果」

Y

生成モデル
(順問題)

音声の線形予測分析とスペクトル最尤推定

■ 3つの観点から解説

■ 「予測誤差」を最小化する観点

■ 最小二乗誤差推定

■ 線形系としての観点

■ 自己回帰系 (AutoRegressive system)

■ 音声の生成過程モデル

■ 最尤推定

■ 白色化

■ スペクトルマッチングとしての観点

■ 最尤スペクトル推定

■ 板倉斎藤距離

「予測誤差」を最小化

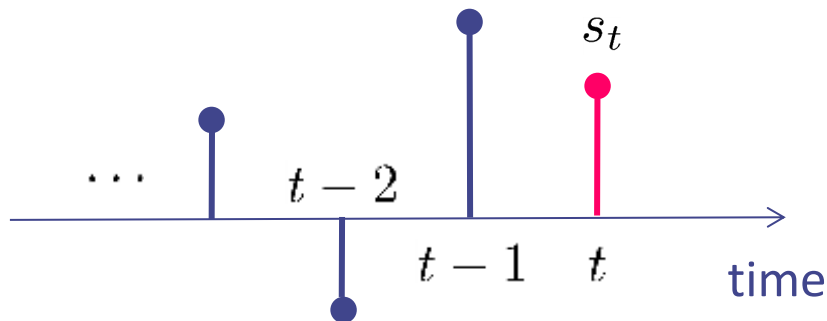
■ 動機: 符号化への応用

- 少ないパラメータで音声信号を表現したい

■ 問題: 線形予測誤差の最小化

- 時刻 t の信号のサンプル値 s_t を、過去のサンプル値 $s_{t-1}, s_{t-2}, \dots, s_{t-P}$ の線形結合で「予測」
- 「予測」の誤差を最小にするには
結合係数(予測係数という)をどう置けば良い？

$$s_t \simeq \sum_{p=1}^P \underline{a_p} s_{t-p}$$



最小二乗誤差推定による定式化

■すべての t で $s_t \simeq \sum_{p=1}^P a_p s_{t-p}$ となる $a_1, \dots, a_P$ を求めたい

■目的関数

$$\mathcal{J}(a_1, \dots, a_P) = \sum_t \left(s_t - \sum_{p=1}^P a_p s_{t-p} \right)^2 \rightarrow \text{minimize}$$

■最小解では $\frac{\partial \mathcal{J}}{\partial a_1} = \frac{\partial \mathcal{J}}{\partial a_2} = \dots = \frac{\partial \mathcal{J}}{\partial a_P} = 0$ を満たすため...

$$\begin{aligned} \longrightarrow \frac{\partial \mathcal{J}}{\partial a_q} &= 2 \sum_t \left(s_t - \sum_{p=1}^P a_p s_{t-p} \right) (-s_{t-q}) \\ &= 2 \left(\sum_p a_p \sum_t s_{t-p} s_{t-q} - \sum_t s_t s_{t-q} \right) \\ &= 0 \quad (q = 1, \dots, P) \end{aligned}$$

最小二乗誤差推定による定式化

■ 連立方程式に帰着

$$\left\{ \begin{array}{l} \sum_p a_p \sum_t s_{t-p} s_{t-1} = \sum_t s_t s_{t-1} \\ \sum_p a_p \sum_t s_{t-p} s_{t-2} = \sum_t s_t s_{t-2} \\ \vdots \\ \sum_p a_p \sum_t s_{t-p} s_{t-P} = \sum_t s_t s_{t-P} \end{array} \right. \longrightarrow c_{q,p} = \sum_t s_{t-p} s_{t-q}$$

以上より最適予測係数は
以下の方程式を満たす

$$\begin{bmatrix} c_{1,1} & \cdots & c_{1,P} \\ \vdots & \ddots & \vdots \\ c_{P,1} & \cdots & c_{P,P} \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_P \end{bmatrix} = \begin{bmatrix} c_{0,1} \\ \vdots \\ c_{0,P} \end{bmatrix}$$

この方程式をYule-Walker方程式という

Levinson-Durbin-Itakuraアルゴリズム

■ Yule-Walker方程式

$$\begin{bmatrix} c_{1,1} & \cdots & c_{1,P} \\ \vdots & \ddots & \vdots \\ c_{P,1} & \cdots & c_{P,P} \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_P \end{bmatrix} = \begin{bmatrix} c_{0,1} \\ \vdots \\ c_{0,P} \end{bmatrix}$$

$$c_{p,q} = v_{|p-q|}$$

$$\begin{bmatrix} v_0 & v_1 & \cdots & v_{P-1} \\ v_1 & v_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & v_1 \\ v_{P-1} & \cdots & v_1 & v_0 \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_P \end{bmatrix} = \begin{bmatrix} v_1 \\ \vdots \\ v_P \end{bmatrix}$$

Toeplitz行列

右辺と左辺
の関係が特殊

連立一次方程式の解き方

■ 一般の場合:

Gaussの消去法

■ 左辺が正値対称行列の場合:

Cholesky分解

■ 左辺がToeplitz行列の場合:

Levinsonアルゴリズム

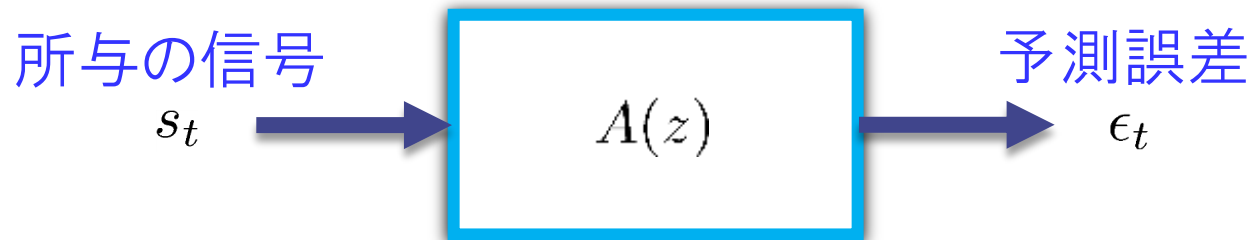
※上記のどれになるかは後述

これを解く高速な解法
(LDIアルゴリズム)が存在

線形系としての解釈

- 所与の信号から予測誤差を出力する線形システム

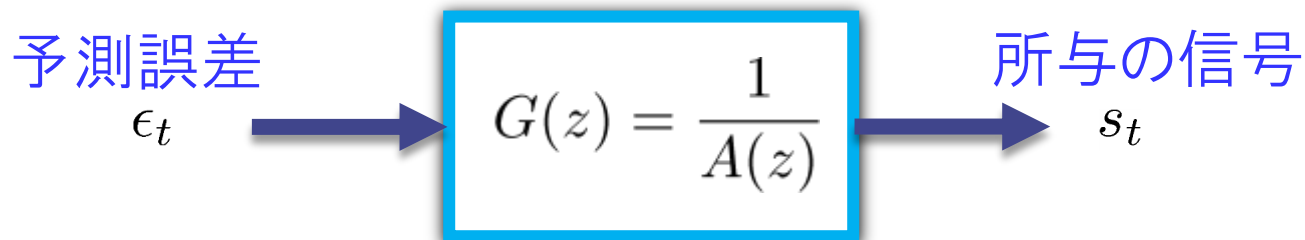
$$\epsilon_t = s_t - \sum_{p=1}^P a_p s_{t-p}$$



$$A(z) = 1 - a_1 z^{-1} - a_2 z^{-2} - \dots - a_P z^{-P}$$

移動平均システム(全零モデル)

- 予測誤差を入力として所与の信号を出力する線形システムは？

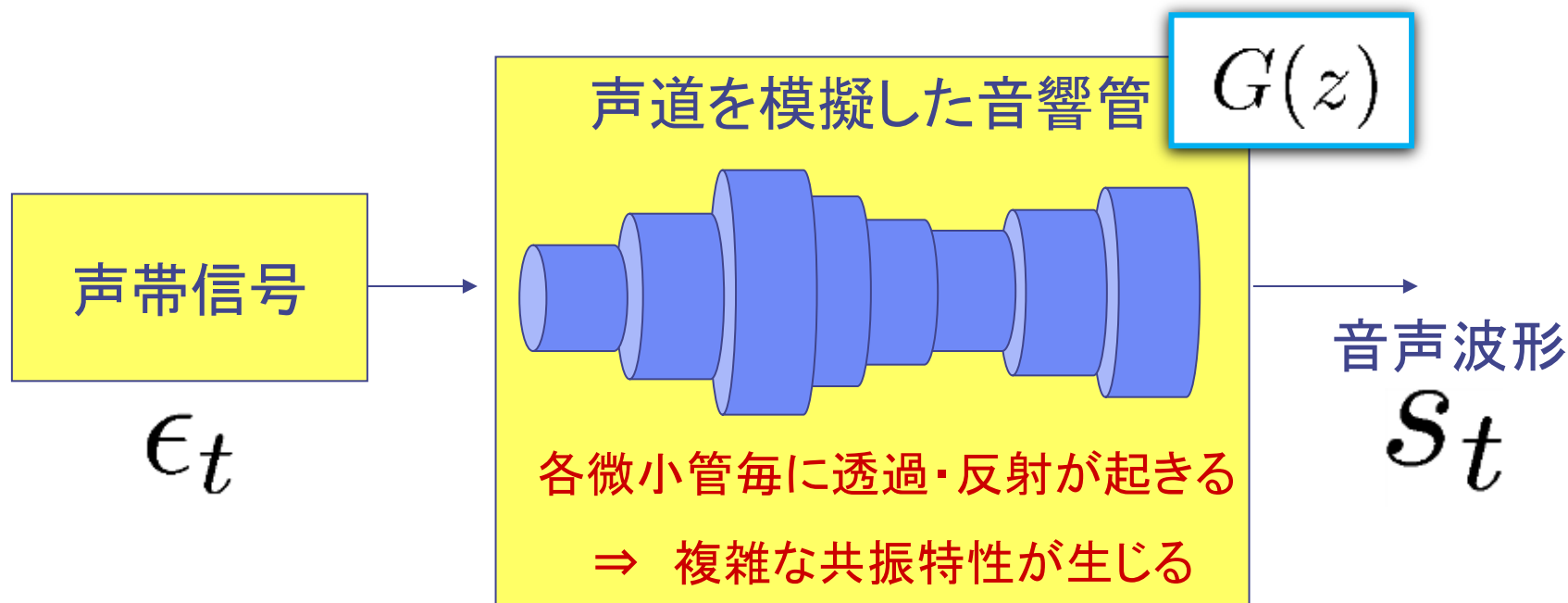


自己回帰システム(全極モデル)

音声生成に適したモデルとは？

■人間の音声生成モデル

- 声帯での基本振動を声道で音色付ける。
- 声道
 - 位置によって太さの異なる音響管の連続と見なせる。
 - 音響管における共振現象 ⇒ 自己回帰(AR)過程



統計モデルによる音声生成過程の表現

■ 声帯振動に関する仮定

- Gauss性 $\dots \epsilon = (\epsilon_1, \dots, \epsilon_T)^T \sim \mathcal{N}(\mu, \Sigma)$
 - 定常性 $\dots \mu = \mathbf{0}, \Sigma : \text{Toeplitz 行列}$
 - 白色性 $\dots \Sigma = \sigma^2 \mathbf{I}$
- } $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$
($t = 1, \dots, T$)

■ 声道特性に関する仮定

- 自己回帰システム(全極モデル)

$$\underbrace{\begin{bmatrix} 1 & & & & 0 \\ -a_1 & 1 & & & \\ \vdots & \ddots & \ddots & & \\ -a_P & & \ddots & \ddots & \\ 0 & & & -a_P & \cdots & -a_1 & 1 \end{bmatrix}}_{\Psi} \underbrace{\begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_T \end{bmatrix}}_{\mathbf{s}} = \underbrace{\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_T \end{bmatrix}}_{\boldsymbol{\epsilon}}$$

最尤推定

■ 今までの仮定をまとめると...

- $\epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$

- $\Psi \mathbf{s} = \epsilon$

未知パラメータは $\{a_p\}_{1 \leq p \leq P}, \sigma^2$

観測されるのは $\{s_t\}_{1 \leq t \leq T}$

■ 観測信号 $\mathbf{s}$ の確率密度関数(尤度関数という)

- $\mathbf{s} = \Psi^{-1} \epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \Psi^{-1} (\Psi^{-1})^T)$

- 対数尤度は

$$-\frac{T}{2} \log(2\pi) - \frac{1}{2} \log |\sigma^2 \Psi^{-1} (\Psi^{-1})^T| - \frac{1}{2\sigma^2} \underbrace{\mathbf{s}^T \Psi^T \Psi \mathbf{s}}_{\rightarrow \|\epsilon\|_2^2} =$$

- logdet項:
$$\begin{aligned} & \frac{\log |\sigma^2 \Psi^{-1} (\Psi^{-1})^T|}{=} \log |\sigma^2 \mathbf{I}| - 2 \log |\Psi| \\ & = T \log \sigma^2 \quad \because |\Psi| = 1 \end{aligned}$$

$$\sum_t \left(s_t - \sum_{p=1}^P a_p s_{t-p} \right)^2$$

白色化効果

- 以上の統計モデルでは ϵ について白色性を仮定していたので、先の最尤推定では ϵ ができるだけ白色になるように $\{a_p\}_{1 \leq p \leq P}$ を決めようとしていたことになる
- このことをよりイメージしやすくするため、以上のモデルを周波数領域で定式化してみよう

周波数領域での定式化

■ 時間領域では・・・

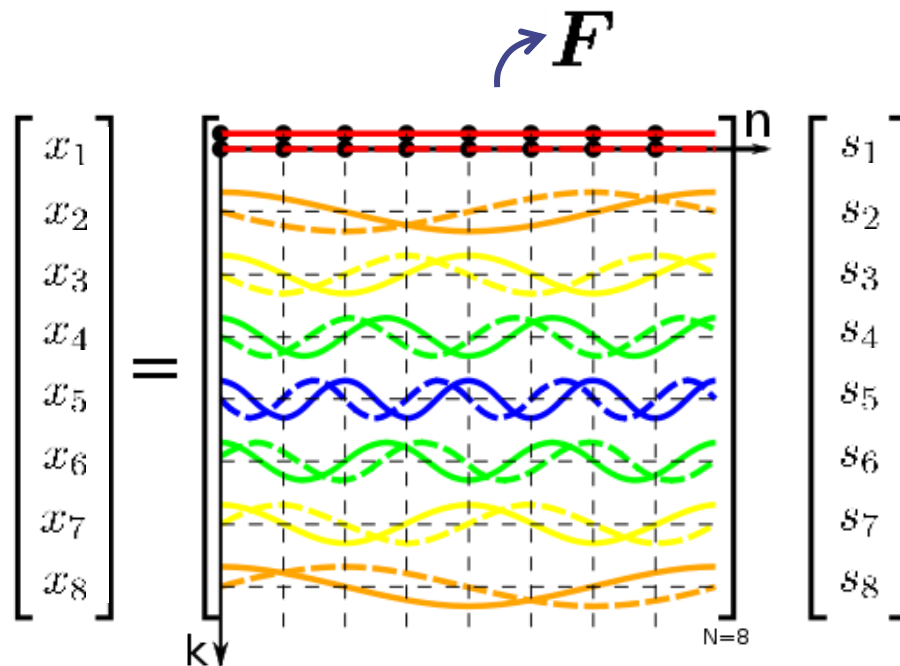
$$\left. \begin{array}{l} \blacksquare \Psi^{-1} \epsilon = s \\ \blacksquare \epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}) \end{array} \right\} \rightarrow s \sim \mathcal{N}(\mathbf{0}, \sigma^2 \Psi^{-1} (\Psi^{-1})^T)$$

■ 周波数領域 (Fourier変換領域) では・・・

■ $x = F s$ (F は離散Fourier変換行列) の確率密度関数は？

$$x \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \sigma^2 \underline{F \Psi^{-1} (\Psi^{-1})^T F^H})$$

$$\begin{aligned} & F \Psi^{-1} (\Psi^{-1})^T F^H \\ &= (F \underline{\Psi^T \Psi} F^H)^{-1} \end{aligned}$$



$\Psi^T \Psi$ について

■ Ψ に関してここでは以下の巡回行列型を仮定

$$\Psi = \begin{bmatrix} 1 & & & -a_P & \cdots & -a_1 \\ -a_1 & 1 & & & \ddots & \vdots \\ \vdots & \ddots & \ddots & & & -a_P \\ -a_P & & \ddots & \ddots & \ddots & \\ 0 & \ddots & & \ddots & \ddots & \\ & & -a_P & \cdots & -a_1 & 1 \end{bmatrix}$$

■ よって $\Psi^T \Psi$ も巡回行列

→ $\Psi^T \Psi$ は離散Fourier変換行列 F によって対角化される

$F \Psi^T \Psi F^H$: 対角行列

$$\sigma^2(F \Psi^T \Psi F^H)^{-1} = \text{diag}(\lambda_1, \dots, \lambda_T)$$

$$\lambda_t = \frac{\sigma^2}{|A(e^{2\pi j(t-1)/T})|^2}$$

$$A(z) = 1 - a_1 z^{-1} - \cdots - a_P z^{-P}$$

周波数成分の確率密度関数

- 以上をまとめると...

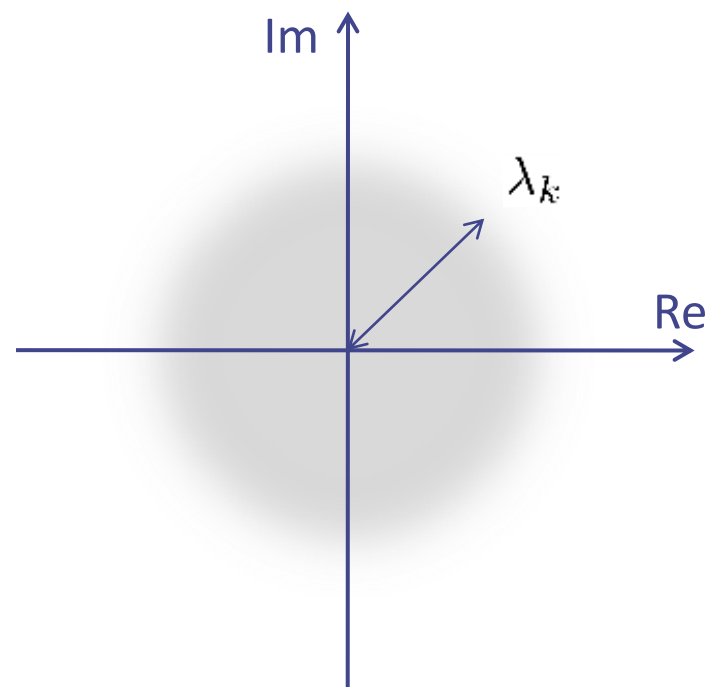
$$\mathbf{x} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \text{diag}(\lambda_1, \dots, \lambda_T))$$

$$\lambda_t = \frac{\sigma^2}{|A(e^{2\pi j(t-1)/T})|^2}$$

$$A(z) = 1 - a_1 z^{-1} - \dots - a_P z^{-P}$$

- 周波数 $2\pi(k-1)/T$ の成分 x_k
 - 他の周波数の成分と独立
 - 分散が λ_k の複素正規分布に従う

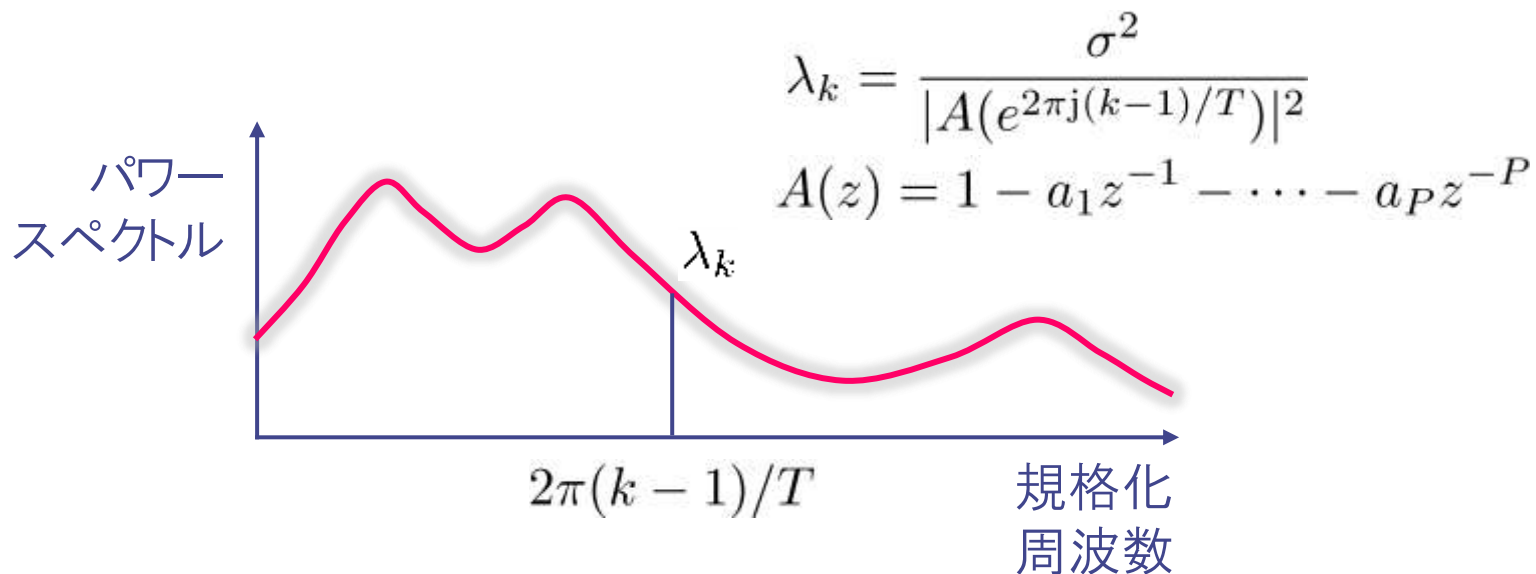
$$x_k \sim \mathcal{N}_{\mathbb{C}}(0, \lambda_k)$$



「スペクトルマッチング」としての見方

- 周波数成分 $\{x_k\}_{1 \leq k \leq T}$ が与えられた下での対数尤度

$$\mathcal{L}(\lambda_1, \dots, \lambda_T) = - \sum_k \left(\log \pi \lambda_k + \frac{|x_k|^2}{\lambda_k} \right)$$



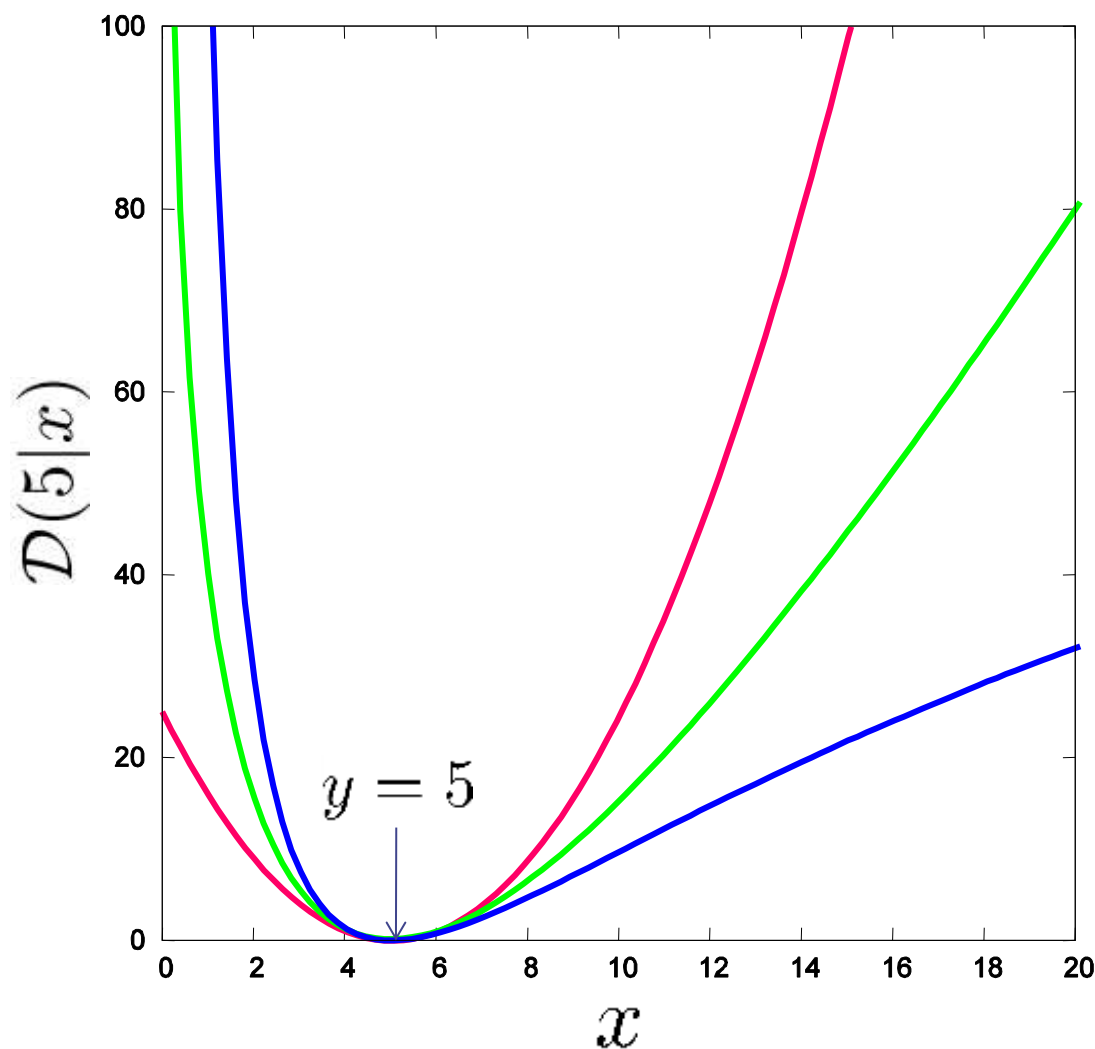
- 上記の対数尤度は、定数項を除けば以下と等しい

$$\mathcal{L}(\lambda_1, \dots, \lambda_T) \stackrel{c}{=} - \sum_k \left(\frac{|x_k|^2}{\lambda_k} - \log \frac{|x_k|^2}{\lambda_k} - 1 \right)$$

板倉 斎藤 距離

板倉斎藤擬距離

■他の擬距離尺度との比較



— 二乗誤差

$$\mathcal{D}(x|y) = (y - x)^2$$

— 1ダイバージェンス

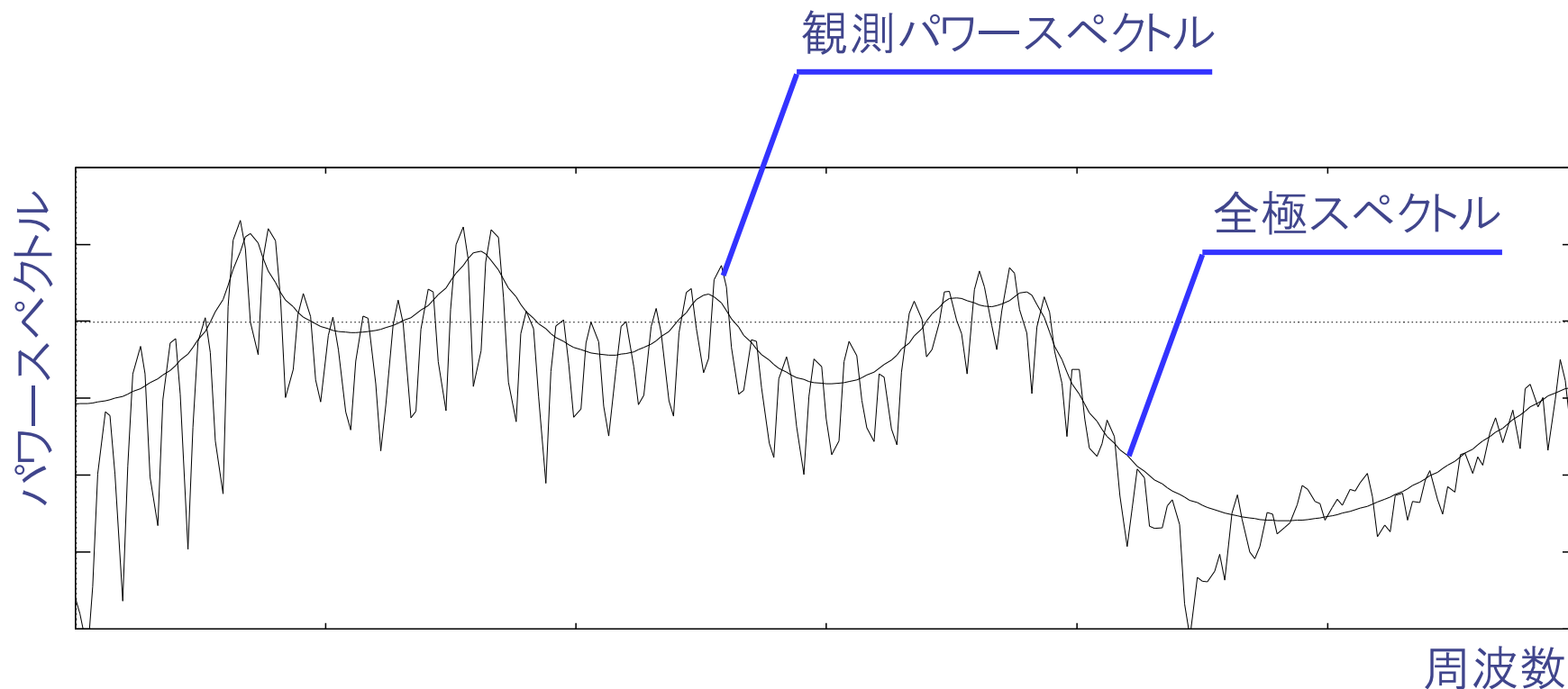
$$\mathcal{D}(x|y) = y \log \frac{y}{x} - (y - x)$$

— 板倉斎藤擬距離

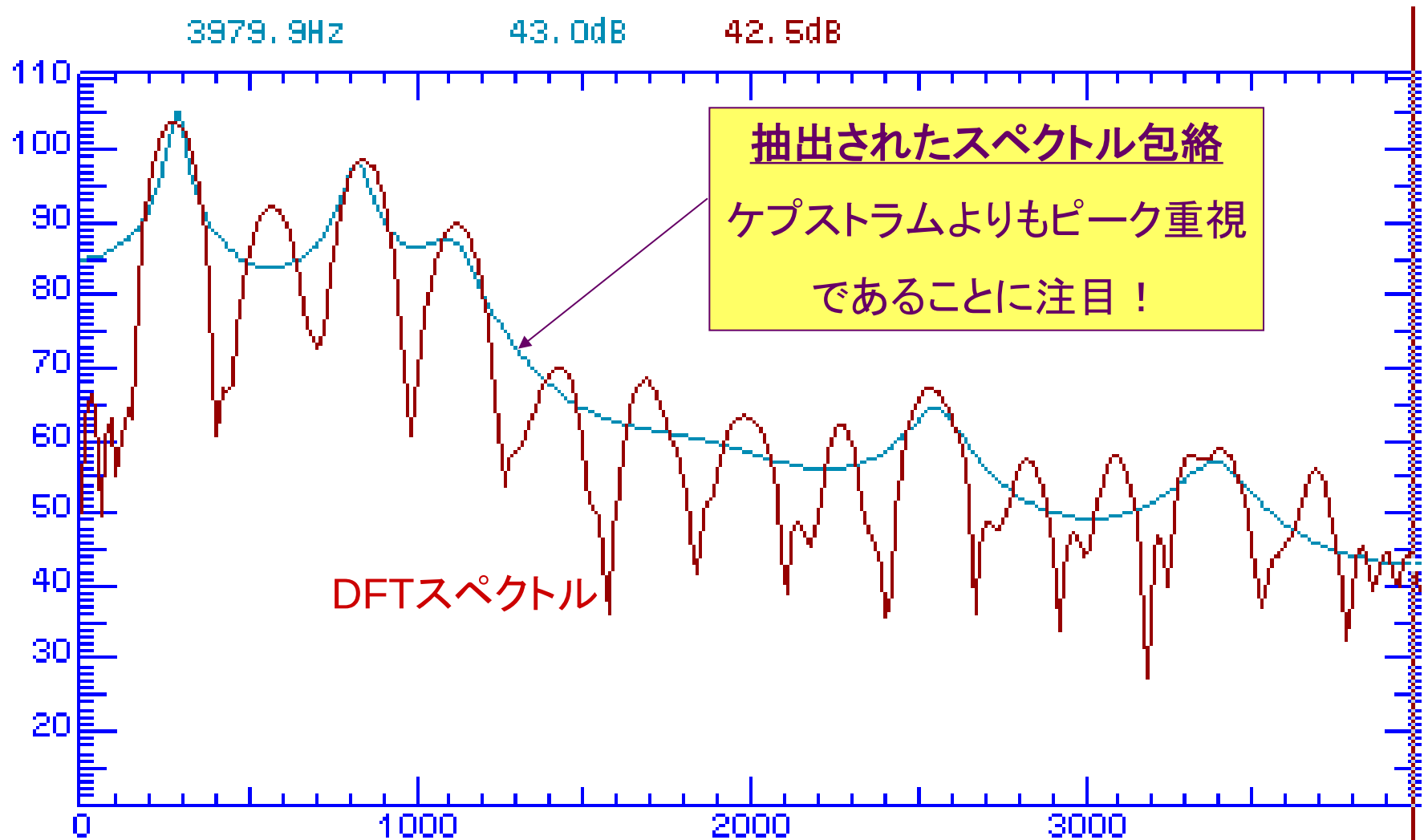
$$\mathcal{D}(x|y) = \frac{y}{x} - \log \frac{y}{x} - 1$$

線形予測分析は「スペクトル包絡」の推定に相当

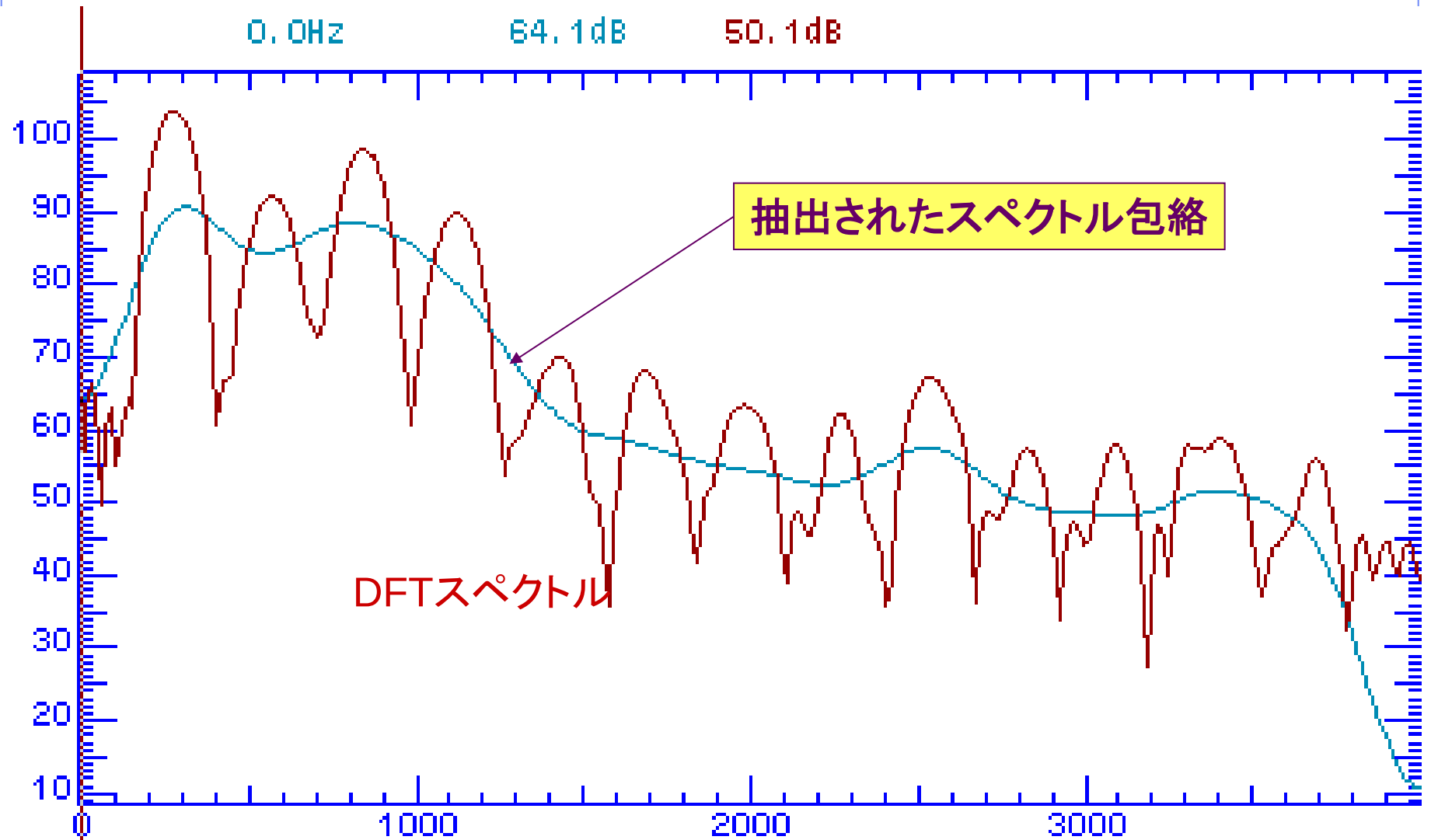
■ 観測パワースペクトルと全極スペクトルとの板倉斎藤距離最小化



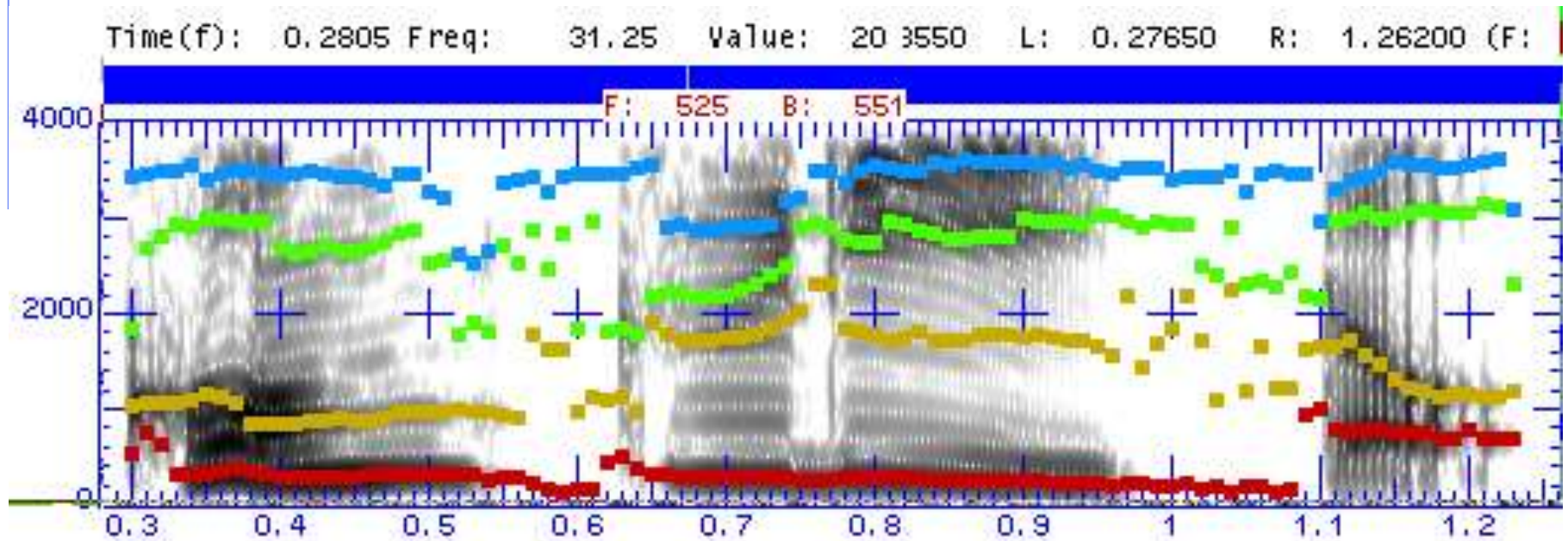
線形予測によるスペクトル包絡の良さは？



(参考) ケプストラムによるスペクトル包絡



線形予測によるホルマント抽出



余談：線形予測法と日本人の貢献

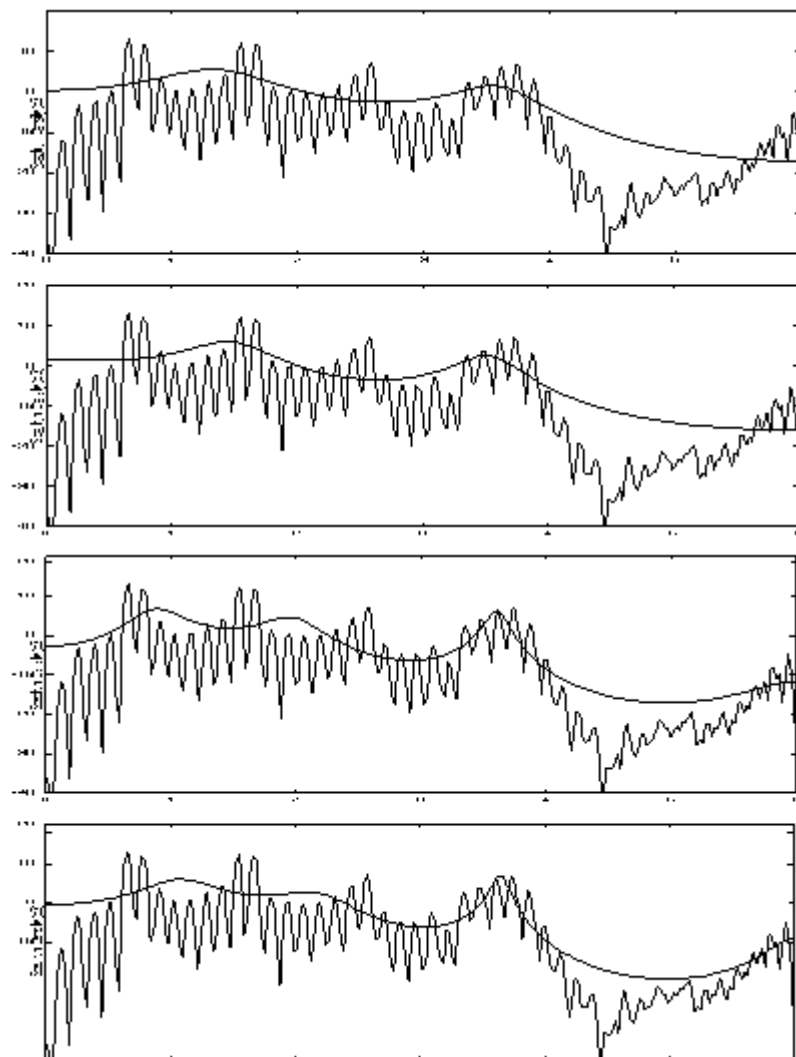
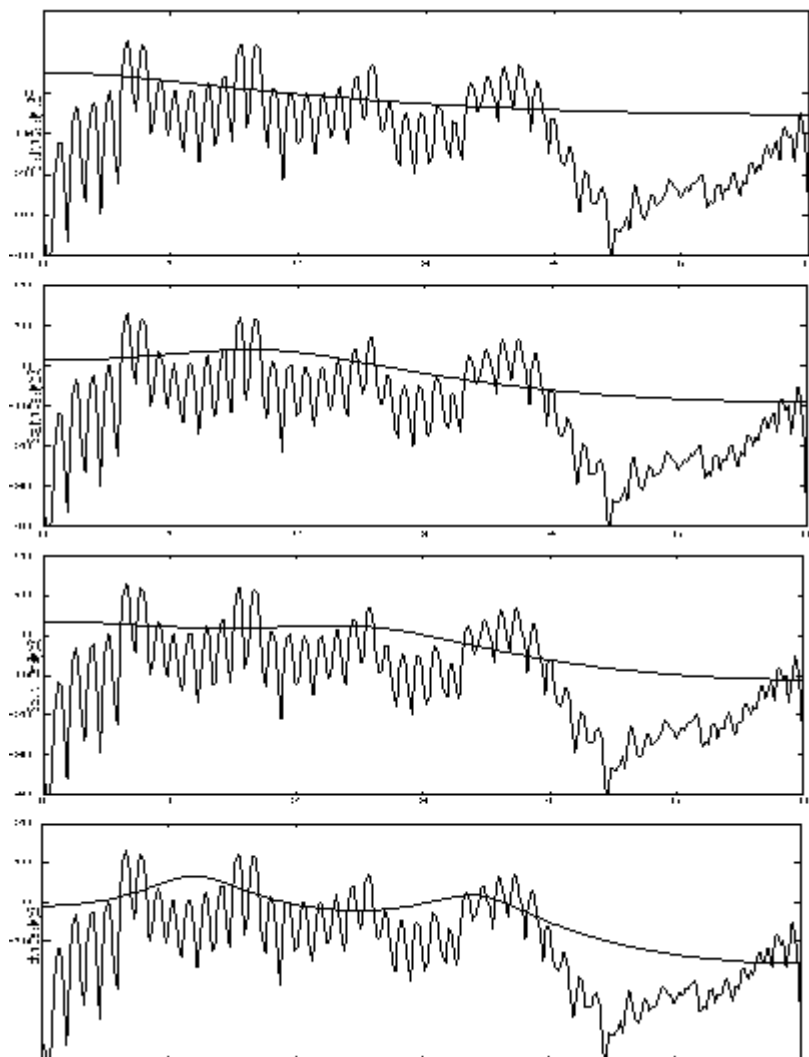
■日経産業新聞 1999年4月20日掲載

「音声認識の研究に金を出すことは価値ある投資だろうか。...音声認識の研究とはまさに「錬金術」に等しい——。」これは、1969年に米国音響学会誌に掲載された寄書の一部である。著者はジョン・ピアス、当時、ベル電話研究所情報通信部門の責任者であった。これを機に、**ピアス傘下にあった音声研究部門では、音声認識の研究が全面的に中止された。**

トップの確信に満ちた判断で中止された研究をボトムアップで再開するのは容易なことではない。**この再開の主役として登場したのは1人の日本人であった。線形予測理論で世界的脚光を浴びていた現NTTの板倉文忠(名古屋大名誉教授)を客員研究員として招いた。**これがベル研での音声認識研究の再始動をうながしたのである。当時、ベル研の研究室長であり、板倉を招いたジェームス・フラナガン(現米国ラトガース大学副学長)は振り返る。**「とても、正面切って音声認識の研究を行える状況ではなかった。部外者である客員研究員が自主的に研究を始めるという苦肉の策を講じ、これが図に当たった」**

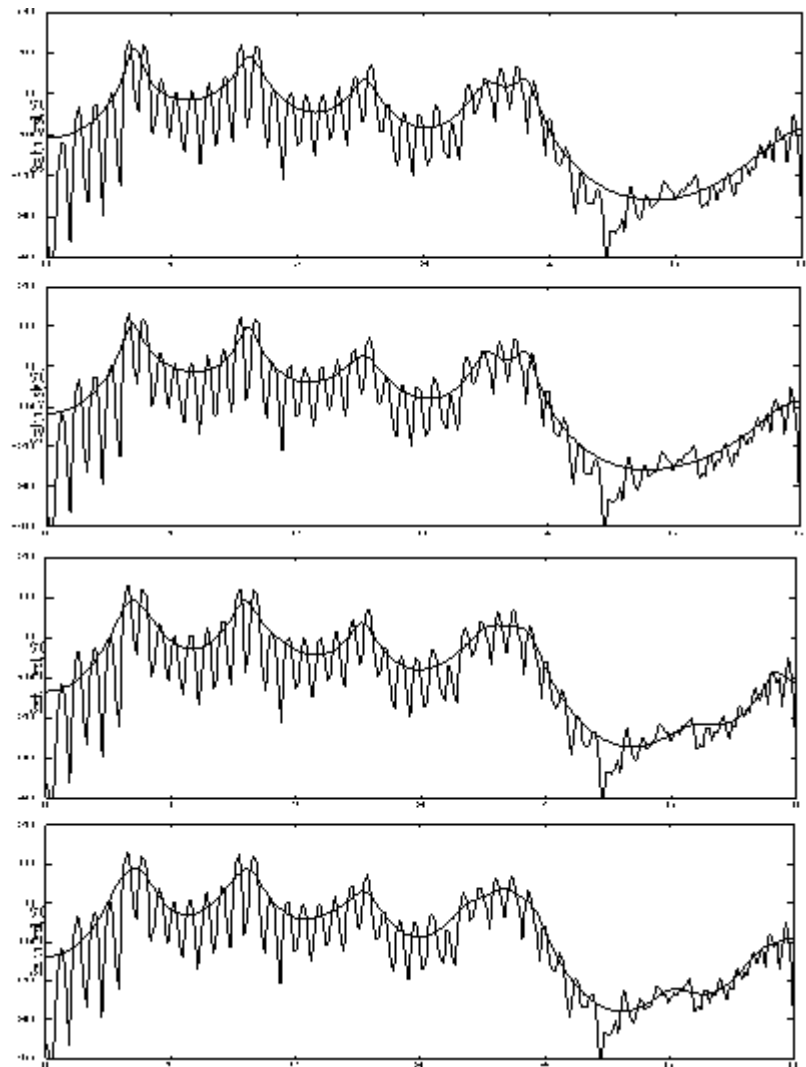
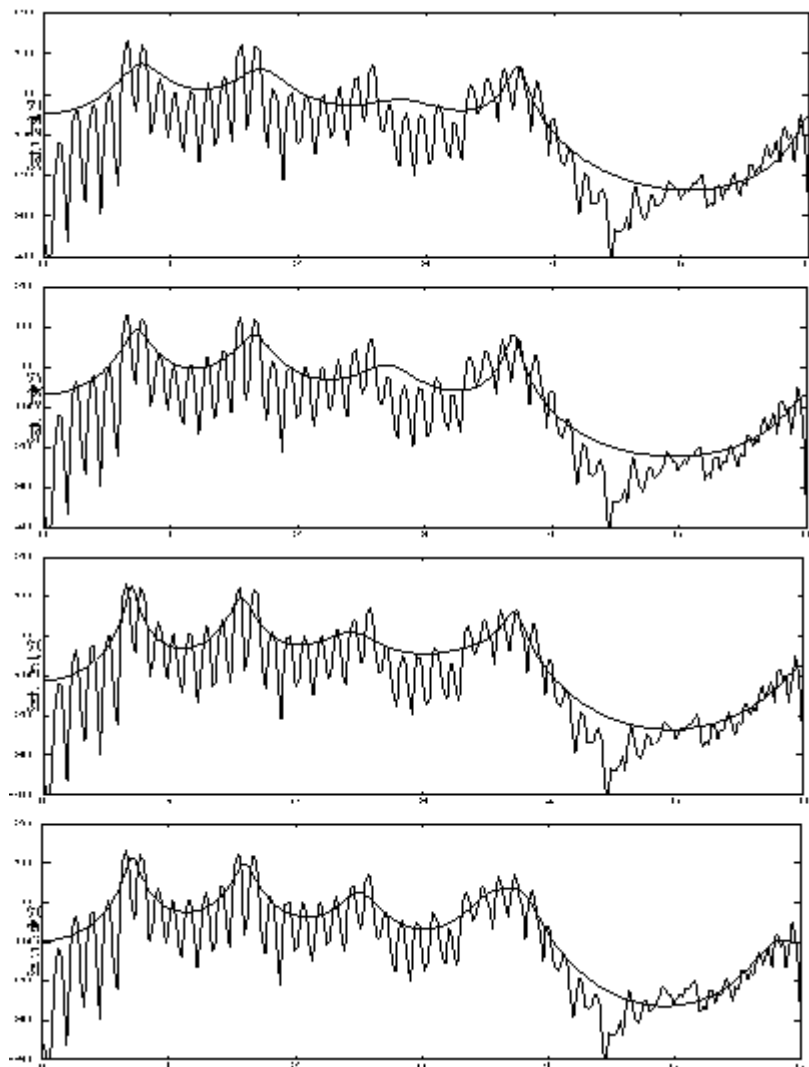
最尤推定による音声スペクトル推定の例

■ 1～8次



最尤推定による音声スペクトル推定の例

■ 9, 10, 12, 14, 16, 18, 22, 26次



Wienerフィルタ

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 $\hat{X}_{\text{MMSE}}(\omega) = \mathbb{E}[X(\omega)|Y(\omega)]$

多変量Gauss分布の性質

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

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

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

■ 線形推定器

$$\hat{s}(t) = \int_{-\infty}^{\infty} h(t - \tau)x(\tau)d\tau = \int_{-\infty}^{\infty} h(\tau)x(t - \tau)d\tau$$

- 観測データ $x(t - \tau)$ の線形結合で推定信号をモデル化

■ 平均二乗誤差最小規範

$$J[h(t)] = \mathbb{E}[|\tilde{s}(t) - s(t)|^2]$$

- J を最小にする $h(t)$ を求めることがここでの問題

Wiener-Hopf積分方程式の解法

復習

- $h(t)$ が非因果的なフィルタの場合

$$R_{sx}(\tau) = \int_{-\infty}^{\infty} R_{xx}(\tau - \alpha)h(\alpha)d\alpha$$

両辺をFourier変換

$$S_{sx}(\omega) = S_{xx}(\omega)H(\omega) \longrightarrow H(\omega) = \frac{S_{sx}(\omega)}{S_{xx}(\omega)}$$

- $R_{sn}(\tau) = 0$ の場合

$$R_{sx}(\tau) = R_{ss}(\tau), \quad R_{xx}(\tau) = R_{ss}(\tau) + R_{nn}(\tau)$$

$$H(\omega) = \frac{S_{ss}(\omega)}{S_{ss}(\omega) + S_{nn}(\omega)}$$

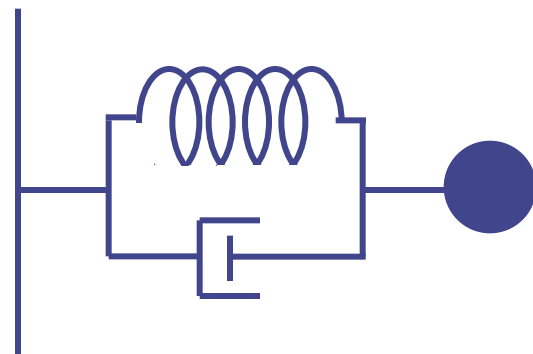
非因果的Wienerフィルタ

多変量ガウス分布の性質から
導きだされる結果と比較せよ

カルマンフィルタ

測定対象に対するモデルの導入

- 例) バネマスダンパ系の質点位置の推定
 - バネマスダンパ系の質点が、ランダムな外力 $F(t)$ により駆動されている
 - 質点の位置は、観測雑音を含む測定器によって観測される



$$m\ddot{x}(t) + \gamma\dot{x}(t) + kx(t) = F(t)$$

$$y(t) = x(t) + \underline{w(t)}$$

観測雑音

離散時間Kalmanフィルタの問題設定

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

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

■ 仮定

k : 時刻インデックス

- v_k, w_k は互いに独立な正規白色雑音

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

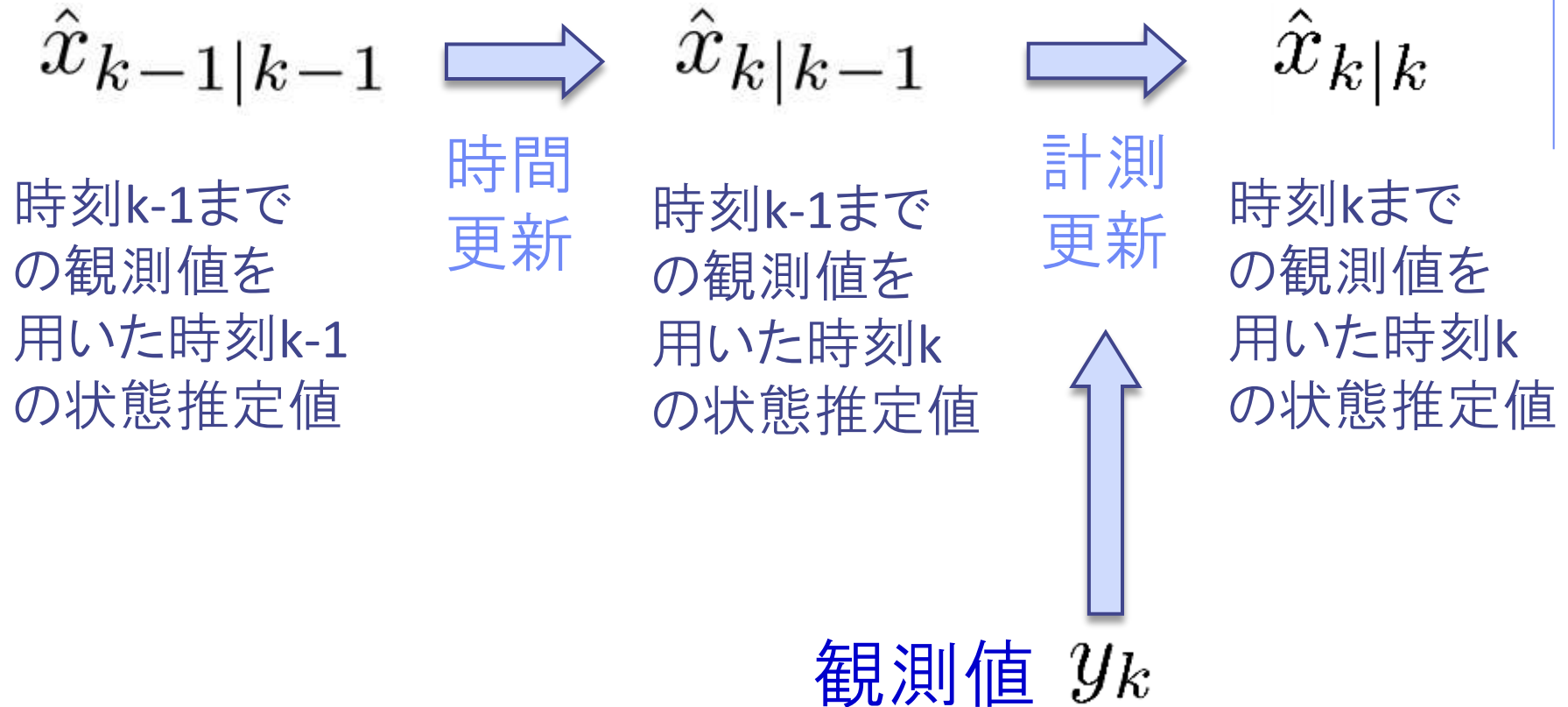
$$\mathbb{E}[w_k] = 0 \quad \mathbb{E}[w_k w_n^T] = W_k \delta_{kn}$$

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

- 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

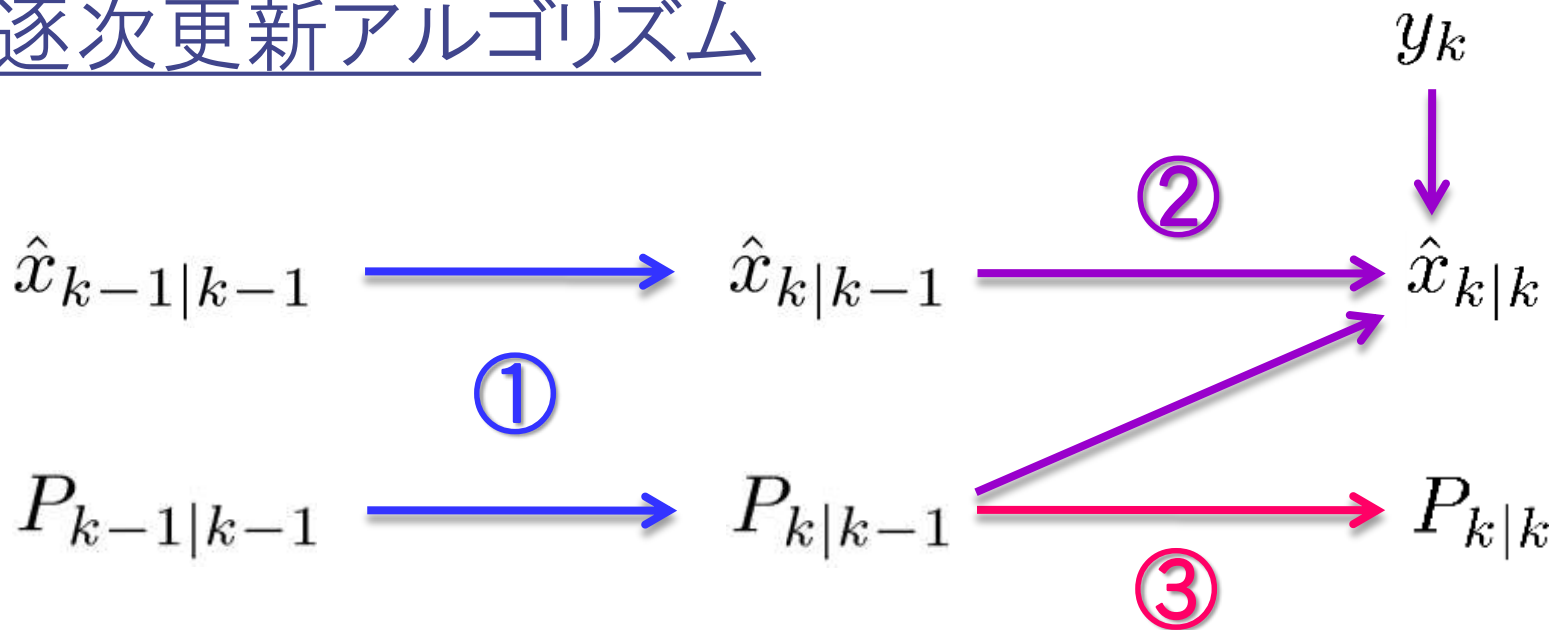


離散時間Kalmanフィルタの目的と導出方針

■ 目的

- $\hat{x}_{k-1|k-1}, P_{k-1|k-1}$ と y_k から $\hat{x}_{k|k}, P_{k|k}$ を逐次的に計算したい
- ただし $P_{n|m} = \text{COV}(x_n - \hat{x}_{n|m})$

逐次更新アルゴリズム



①状態の時間更新

導出の詳細は信号処理論2講義資料参照

$$\hat{x}_{k-1|k-1} \longrightarrow \hat{x}_{k|k-1}$$

$$\hat{x}_{k|k-1} = \Phi_{k,k-1} \hat{x}_{k-1|k-1}$$

$$P_{k-1|k-1} \longrightarrow P_{k|k-1}$$

$$\begin{aligned} P_{k|k-1} &= \text{COV}(x_k - \hat{x}_{k|k-1}) \\ &= \mathbb{E}[(x_k - \hat{x}_{k|k-1})(x_k - \hat{x}_{k|k-1})^T] \\ &= \Phi_{k,k-1} P_{k-1|k-1} \Phi_{k,k-1}^T + \underline{B_k V_k B_k^T} \end{aligned}$$

時間更新による分散の増分

②状態の計測更新

■ $\hat{x}_{k|k-1}, y_k \longrightarrow \hat{x}_{k|k}$

■ イノベーション(観測値に対する予測の誤差)

$$\mu_k = y_k - C_k \hat{x}_{k|k-1}$$



■ 更新式

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + \underline{K_k} \mu_k$$

Kalmanゲイン(未知)

(参考) (θ, y) がガウス分布に従うとき

$$\theta_{\text{MMSE}} = \mathbb{E}[\theta|y] = \bar{\theta} + A_0(y - \bar{y})$$

③状態推定値の誤差共分散の更新

$$\blacksquare P_{k|k-1} \longrightarrow P_{k|k}$$

$$P_{k|k}$$

$$= \text{cov}(x_k - \hat{x}_{k|k})$$

$$= \text{cov}(x_k - (\hat{x}_{k|k-1} + K_k \underline{\mu_k}))$$

$$= \text{cov}(x_k - (\hat{x}_{k|k-1} + K_k (\underline{y_k} - C_k \hat{x}_{k|k-1})))$$

$$= \text{cov}(\underline{x_k} - (\underline{\hat{x}_{k|k-1}} + K_k (C_k \underline{x_k} + w_k - C_k \underline{\hat{x}_{k|k-1}})))$$

$$= \text{cov}((I - K_k C_k)(x_k - \hat{x}_{k|k-1}) - K_k w_k) \quad \because w \text{ は } x, \hat{x} \text{ と独立}$$

$$= \text{cov}((I - K_k C_k)(x_k - \hat{x}_{k|k-1})) + \text{cov}(K_k w_k)$$

$$= (I - K_k C_k) \text{cov}(x_k - \hat{x}_{k|k-1}) (I - K_k C_k)^T + K_k \text{cov}(w_k) K_k^T$$

$$= (I - K_k C_k) P_{k|k-1} (I - K_k C_k)^T + K_k W_k K_k^T$$

②最適Kalmanゲインの導出

■ Kalmanゲインを決定する最適化問題

$$\mathbb{E}[\|x_k - \hat{x}_{k|k}\|_2^2] \rightarrow \text{minimize}$$

$$\begin{aligned}\mathbb{E}[\|x_k - \hat{x}_{k|k}\|_2^2] &= \mathbb{E}[(x_k - \hat{x}_{k|k})^\top (x_k - \hat{x}_{k|k})] \\ &= \text{tr}(\mathbb{E}[(x_k - \hat{x}_{k|k})(x_k - \hat{x}_{k|k})^\top]) \\ &= \text{tr}(\underline{P_{k|k}}) \\ &= \text{tr}((I - K_k C_k) P_{k|k-1} (I - K_k C_k)^\top + K_k W_k K_k^\top)\end{aligned}$$

②最適Kalmanゲインの導出

■最適化規準

$$\begin{aligned}\text{tr}(P_{k|k}) &= \text{tr}((I - K_k C_k) P_{k|k-1} (I - K_k C_k)^\top + K_k W_k K_k^\top) \\ &= \text{tr}(P_{k|k-1} - \underline{K_k} C_k P_{k|k-1} - P_{k|k-1} C_k^\top \underline{K_k}^\top \\ &\quad + \underline{K_k} C_k P_{k|k-1} C_k^\top \underline{K_k}^\top + \underline{K_k} W_k \underline{K_k}^\top)\end{aligned}$$

■ $\text{tr}(P_{k|k})$ を最小化する K_k は $\frac{\partial \text{tr}(P_{k|k})}{\partial K_k} = 0$ を満たす

$$\begin{aligned}\frac{\partial \text{tr}(P_{k|k})}{\partial K_k} &= -2P_{k|k-1} C_k^\top + 2K_k (C_k P_{k|k-1} C_k^\top + W_k) \\ &= 0\end{aligned}$$

∵トレースの微分公式より

$$\frac{\partial \text{tr}(MK^\top)}{\partial K} = M, \quad \frac{\partial \text{tr}(KMK^\top)}{\partial K} = KM^\top + KM$$

$$\therefore \hat{K}_k = P_{k|k-1} C_k^\top (C_k P_{k|k-1} C_k^\top + W_k)^{-1}$$

③最適Kalmanゲインにおける推定誤差共分散

- $P_{k|k}$ に $\hat{K}_k$ を代入

$$P_{k|k} = (I - K_k C_k) P_{k|k-1} (I - K_k C_k)^T + K_k W_k K_k^T$$

$$\hat{K}_k = P_{k|k-1} C_k^T (C_k P_{k|k-1} C_k^T + W_k)^{-1}$$

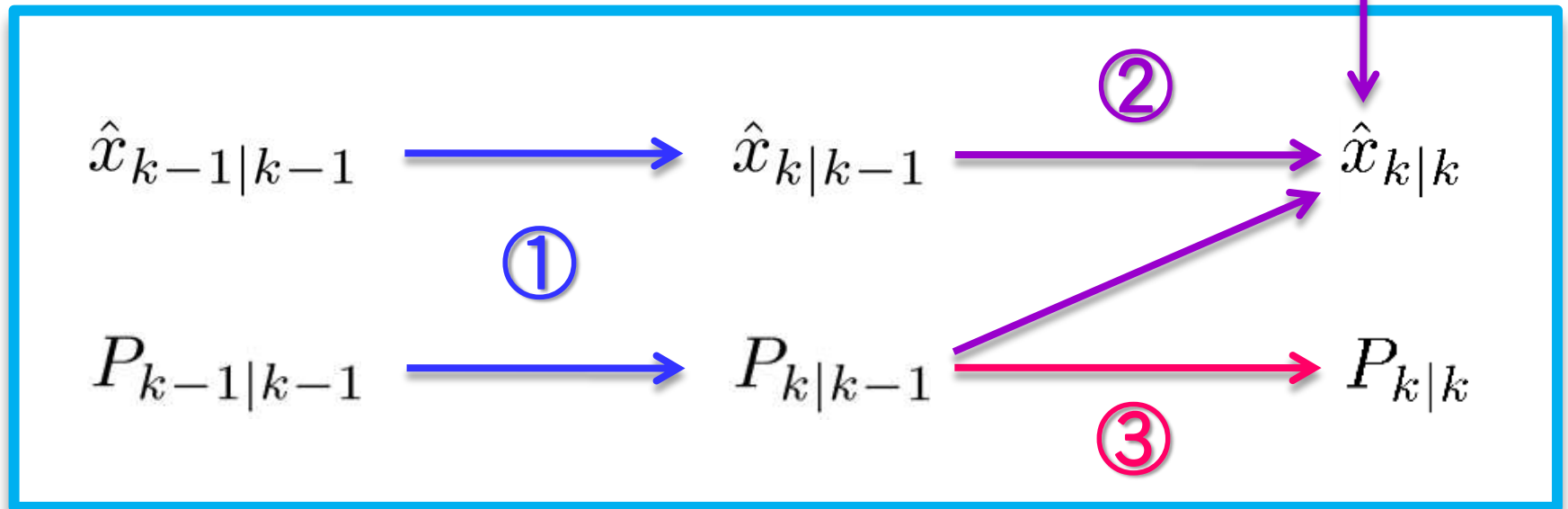
(最適Kalmanゲイン)



$$\hat{K}_k (C_k P_{k|k-1} C_k^T + W_k) \hat{K}_k^T = P_{k|k-1} C_k^T \hat{K}_k^T$$


$$P_{k|k} = (I - \hat{K}_k C_k) P_{k|k-1}$$

離散時間Kalmanフィルタのまとめ



- ①
$$\begin{aligned}\hat{x}_{k|k-1} &= \Phi_{k,k-1} \hat{x}_{k-1|k-1} \\ P_{k|k-1} &= \Phi_{k,k-1} P_{k-1|k-1} \Phi_{k,k-1}^T + B_k V_k B_k^T\end{aligned}$$
- ②
$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + P_{k|k-1} C_k^T (C_k P_{k|k-1} C_k^T + W_k)^{-1} (y_k - C_k \hat{x}_{k|k-1})$$
- ③
$$P_{k|k} = (I - P_{k|k-1} C_k^T (C_k P_{k|k-1} C_k^T + W_k)^{-1} C_k) P_{k|k-1}$$

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

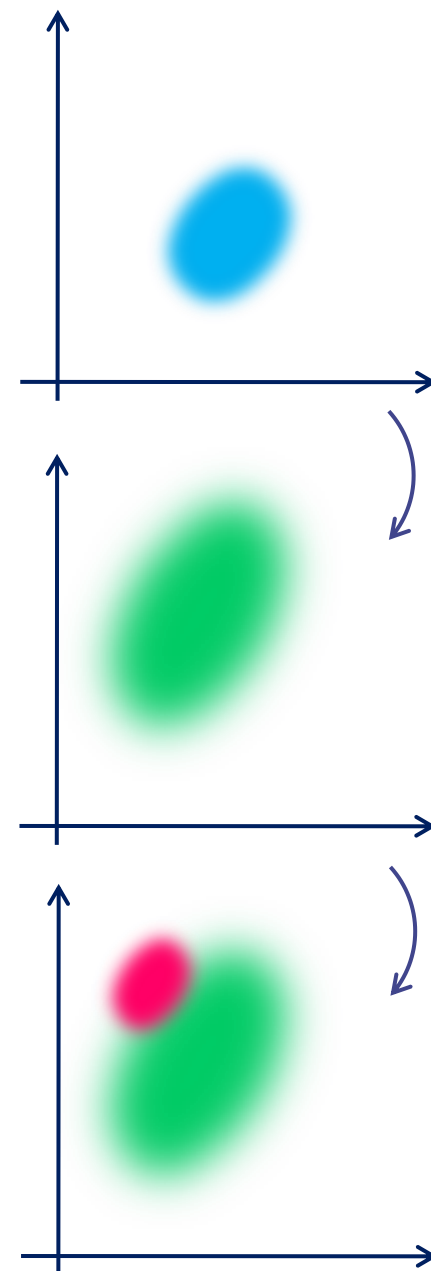
$\hat{x}_{k-1|k-1}, P_{k-1|k-1}$: この分布の平均と共分散行列に相当

■ $p(x_k|y_{1:k-1})$

- 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後確率分布



離散時間KalmanフィルタのBayes的解釈

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

システムモデルより:

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

計測モデルより:

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

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

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

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