

東京大学工学部 4年生 夏学期

応用音響学 第3回 (5/01)

猿渡 洋

東京大学大学院情報理工学系研究科
創造情報学/システム情報学専攻
hiroshi_saruwatari@ipc.i.u-tokyo.ac.jp

2020年度講義スケジュール

前半(猿渡担当)

- 4/17: 第1回
- 4/24: 第2回
- 5/01: 第3回
- 5/08: 第4回
- 5/15: 第5回
- 6/05: 第6回

後半(牧野先生担当)

- 6/12: 第7回
- 6/19: 第8回
- 6/26: 第9回
- 7/03: 第10回
- 7/10: 第11回
- 7/17: 学期末試験(予定)

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

- 学期末試験

ケプストラム処理の特徴

復習

■ スペクトル構造の分解

- 対数を利用してスペクトル積を和に変換
- ケフレンシー領域へ変換することにより、スペクトル包絡と周期的微細構造を区別可能にする。
- 単純な窓かけ操作(リフター)により、包絡成分のみ(もしくは微細構造のみ)を抽出可能

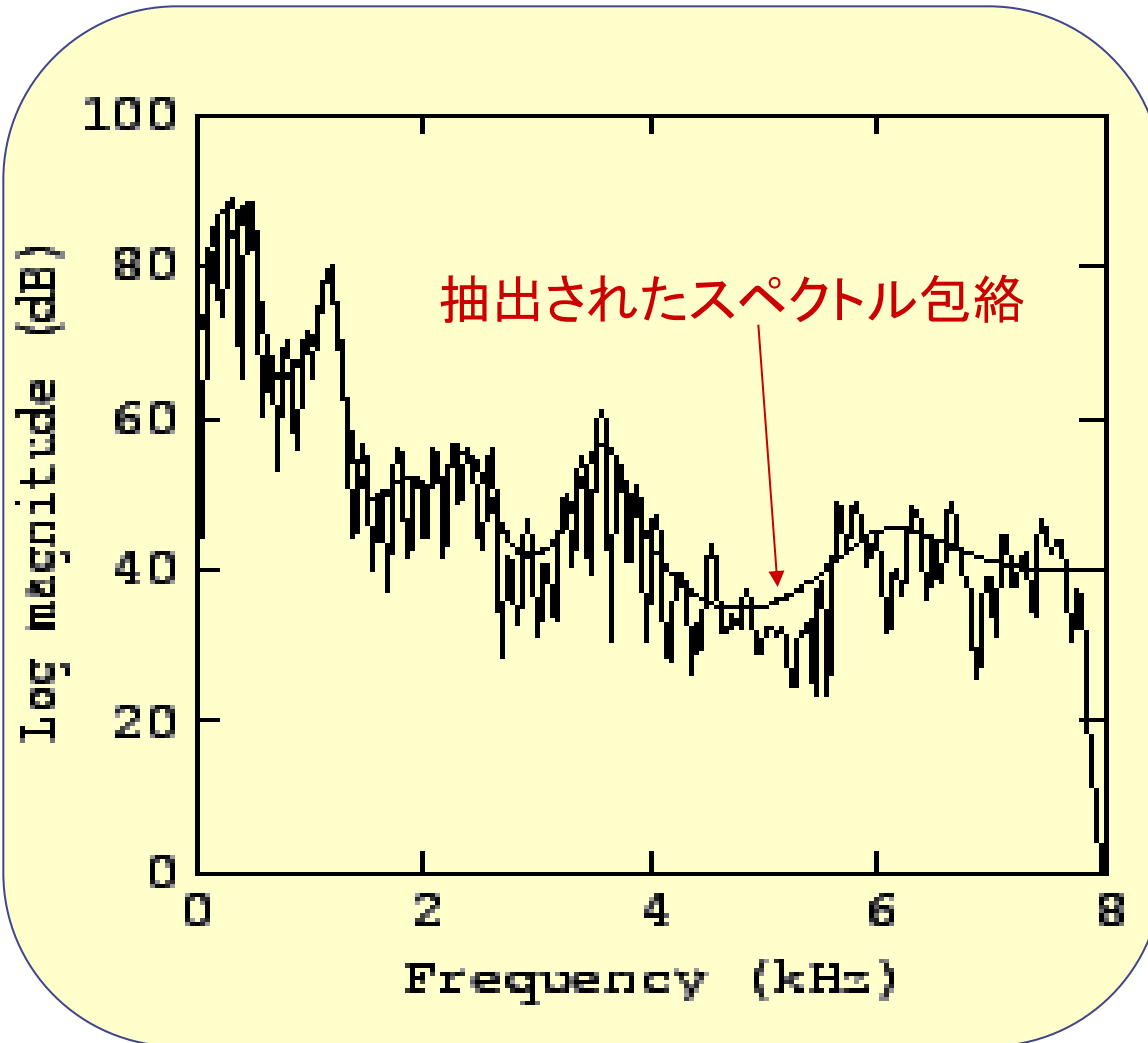
■ 少ない演算量

- スペクトル包絡成分を抽出するのに必要な演算
[対数演算＋逆DFT＋リフタリング＋DFT]
⇒ 非常に少ない演算量で抽出可能

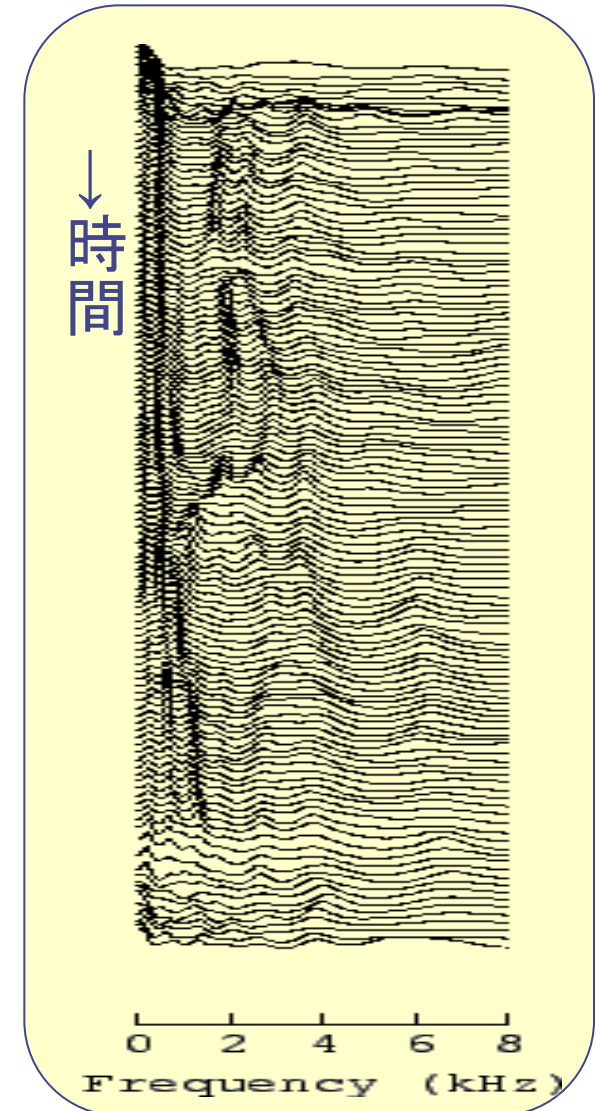
ケプストラムによるスペクトル包絡

復習

短時間スペクトル包絡の例



スペクトル包絡の時間遷移



ケプストラム分析のまとめ

復習

■長所

- 比較的単純な操作でスペクトル包絡抽出可能
- 高次ケプストラムも使用すれば基本周波数も抽出可能

■問題点

- リフタリングのカットオフ位置をどのようにして決めるか？
- 抽出されたスペクトル包絡において、ホルマント共振があまり強く表示されない。

人間の聴覚系では共振点をより聞いていると言われている

⇒声道での共振をモデルにしたパラメトリック分析が有効

本日の話題

■線形予測分析 (Linear Predictive Coding)

- 音声情報処理研究の歴史の幕開けとなった信号処理技術（統計的手法を取り入れた初めての音声研究として有名）

- 音声分析合成(ボコーダ)
 - 音声音響符号化
 - 音声認識のための音声特徴量
 - 音声強調(残響除去、ブラインド音声分離)
- などへの応用

- 日本発の技術としても知られる

- Levinson-Durbin-板倉アルゴリズム、偏自己相関(PARCOR)、線スペクトル対(Line Spectrum Pair) の発明や板倉齋藤距離の発見など、板倉文忠氏(名古屋大学名誉教授)の電電公社時代の活躍が世界的に有名

線形予測分析

■ 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

線形予測分析

■ 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

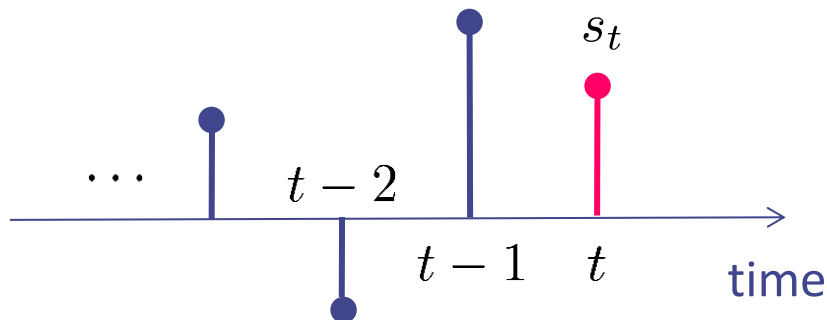
■ 動機: 符号化への応用

- 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 (1/6)

■ 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

この場合の
解き方は？

Levinson-Durbin-Itakuraアルゴリズム (2/6)

■ 右辺を左辺に移項

$$\begin{bmatrix} v_0 & v_1 & \cdots & v_{P-1} \\ v_1 & v_0 & \cdots & v_{P-2} \\ \vdots & \vdots & \ddots & \vdots \\ v_{P-1} & v_{P-2} & \cdots & v_0 \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_P \end{bmatrix} = \begin{bmatrix} v_1 \\ \vdots \\ v_P \end{bmatrix}$$

$P \times P$ 行列

$$\sigma_P^2 = v_0 - \sum_{p=1}^P v_p a_p$$

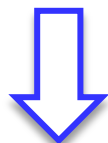

$$\begin{bmatrix} v_0 & v_1 & \cdots & v_P \\ v_1 & v_0 & \cdots & v_{P-1} \\ \vdots & \vdots & \ddots & \vdots \\ v_P & v_{P-1} & \cdots & v_0 \end{bmatrix} \begin{bmatrix} 1 \\ -a_1 \\ \vdots \\ -a_P \end{bmatrix} = \begin{bmatrix} \sigma_P^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$(P+1) \times (P+1)$ 行列

Levinson-Durbin-Itakura アルゴリズム (3/6)

- P 次の (最適な) 予測係数から、(P+1) 次の (最適な) 予測係数を再帰的に解けないか？

$$\begin{bmatrix} v_0 & v_1 & \cdots & v_P \\ v_1 & v_0 & \cdots & v_{P-1} \\ \vdots & \vdots & \ddots & \vdots \\ v_P & v_{P-1} & \cdots & v_0 \end{bmatrix} \begin{bmatrix} 1_{(P)} \\ -a_1^{(P)} \\ \vdots \\ -a_P^{(P)} \end{bmatrix} = \begin{bmatrix} \sigma_P^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \dots (*)$$



$$\begin{bmatrix} v_0 & v_1 & \cdots & v_{P+1} \\ v_1 & v_0 & \cdots & v_P \\ \vdots & \vdots & \ddots & \vdots \\ v_{P+1} & v_P & \cdots & v_0 \end{bmatrix} \begin{bmatrix} 1_{(P+1)} \\ -a_1^{(P+1)} \\ \vdots \\ -a_{P+1}^{(P+1)} \end{bmatrix} = \begin{bmatrix} \sigma_{P+1}^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\hat{a}_1^{(P)}, \dots, \hat{a}_P^{(P)}$$

関係は？

$$\hat{a}_1^{(P+1)}, \dots, \hat{a}_P^{(P+1)}$$

Levinson-Durbin-Itakura アルゴリズム (4/6)

■ 式(*)を変形

$$\begin{bmatrix} v_0 & v_1 & \cdots & v_P \\ v_1 & v_0 & \cdots & v_{P-1} \\ \vdots & \vdots & \ddots & \vdots \\ v_P & v_{P-1} & \cdots & v_0 \end{bmatrix} \begin{bmatrix} 1_{(P)} \\ -a_1^{(P)} \\ \vdots \\ -a_P^{(P)} \end{bmatrix} = \begin{bmatrix} \sigma_P^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

(P+1) × (P+1) 行列

$$\begin{bmatrix} v_0 & v_1 & \cdots & v_P & v_{P+1} \\ v_1 & v_0 & \cdots & v_{P-1} & v_P \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ v_P & v_{P-1} & \cdots & v_0 & v_1 \end{bmatrix} \begin{bmatrix} 1_{(P)} \\ -a_1^{(P)} \\ \vdots \\ -a_P^{(P)} \\ 0 \end{bmatrix} = \begin{bmatrix} \sigma_P^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} v_0 & v_1 & \cdots & v_P & v_{P+1} \\ v_1 & v_0 & \cdots & v_{P-1} & v_P \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ v_P & v_{P-1} & \cdots & v_0 & v_1 \\ v_{P+1} & v_P & \cdots & v_1 & v_0 \end{bmatrix} \begin{bmatrix} 1_{(P)} \\ -a_1^{(P)} \\ \vdots \\ -a_P^{(P)} \\ 0 \end{bmatrix} = \begin{bmatrix} \sigma_P^2 \\ 0 \\ \vdots \\ 0 \\ w_P \end{bmatrix}$$

(P+2) × (P+2) 行列

$$w_P \rightarrow v_{P+1} - \sum_{p=1}^P a_p v_{P-p+1}$$

Levinson-Durbin-Itakuraアルゴリズム (5/6)

■ 左辺は対称行列より

$$\begin{bmatrix} v_0 & v_1 & \cdots & v_P & v_{P+1} \\ v_1 & v_0 & \cdots & v_{P-1} & v_P \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ v_P & v_{P-1} & \cdots & v_0 & v_1 \\ v_{P+1} & v_P & \cdots & v_1 & v_0 \end{bmatrix} \begin{bmatrix} 1 \\ -a_1^{(P)} \\ \vdots \\ -a_P^{(P)} \\ 0 \end{bmatrix} = \begin{bmatrix} \sigma_P^2 \\ 0 \\ \vdots \\ 0 \\ w_P \end{bmatrix} \quad \dots \textcircled{1}$$

$$\begin{bmatrix} v_0 & v_1 & \cdots & v_P & v_{P+1} \\ v_1 & v_0 & \cdots & v_{P-1} & v_P \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ v_P & v_{P-1} & \cdots & v_0 & v_1 \\ v_{P+1} & v_P & \cdots & v_1 & v_0 \end{bmatrix} \begin{bmatrix} 0 \\ -a_P^{(P)} \\ \vdots \\ -a_1^{(P)} \\ 1 \end{bmatrix} = \begin{bmatrix} w_P \\ 0 \\ \vdots \\ 0 \\ \sigma_P^2 \end{bmatrix} \quad \dots \textcircled{2}$$

■ ① - k_P × ② （ k_P は任意の係数）

$$\begin{bmatrix} v_0 & v_1 & \cdots & v_P & v_{P+1} \\ v_1 & v_0 & \cdots & v_{P-1} & v_P \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ v_P & v_{P-1} & \cdots & v_0 & v_1 \\ v_{P+1} & v_P & \cdots & v_1 & v_0 \end{bmatrix} \begin{bmatrix} 1 - 0 \\ -a_1^{(P)} + k_P a_P^{(P)} \\ \vdots \\ -a_P^{(P)} + k_P a_1^{(P)} \\ 0 - k_P \end{bmatrix} = \begin{bmatrix} \sigma_P^2 - k_P w_P \\ 0 \\ \vdots \\ 0 \\ w_P - k_P \sigma_P^2 \end{bmatrix}$$

Levinson-Durbin-Itakura アルゴリズム (6/6)

$$\begin{bmatrix} v_0 & v_1 & \cdots & v_P & v_{P+1} \\ v_1 & v_0 & \cdots & v_{P-1} & v_P \\ \vdots & \vdots & \ddots & \vdots & \\ v_P & v_{P-1} & \cdots & v_0 & v_1 \\ v_{P+1} & v_P & \cdots & v_1 & v_0 \end{bmatrix} \begin{bmatrix} 1 - 0 \\ -a_1^{(P)} + k_P a_P^{(P)} \\ \vdots \\ -a_P^{(P)} + k_P a_1^{(P)} \\ 0 - k_P \end{bmatrix} = \begin{bmatrix} \sigma_P^2 - k_P w_P \\ 0 \\ \vdots \\ 0 \\ w_P - k_P \sigma_P^2 \end{bmatrix}$$

- k_P は任意なので $w_P - k_P \sigma_P^2 = 0$ となるように k_P を選ぶと上式は下記の形になる

$$\begin{bmatrix} v_0 & v_1 & \cdots & v_{P+1} \\ v_1 & v_0 & \cdots & v_P \\ \vdots & \vdots & \ddots & \vdots \\ v_{P+1} & v_P & \cdots & v_0 \end{bmatrix} \begin{bmatrix} 1 \\ -a_1^{(P+1)} \\ \vdots \\ -a_P^{(P+1)} \end{bmatrix} = \begin{bmatrix} \sigma_{P+1}^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

- よって以下の再帰式を得る

$$k_P = \frac{w_P}{\sigma_P^2} \longrightarrow \text{偏自己相関(Partial Correlation; PARCOR)係数}$$

[Itakura1969]

$$\sigma_{P+1}^2 = \sigma_P^2 - k_P w_P$$

$$a_p^{(P+1)} = a_p^{(P)} - k_P a_{P-p+1}^{(P)}$$

なお、明らかに $a_1^{(1)} = \frac{v_1}{v_0}$

予測誤差

- 最適予測係数を $\hat{a}_1, \dots, \hat{a}_P$ とすると、「予測の誤差」は...

$$\epsilon_t = s_t - \sum_{p=1}^P \hat{a}_p s_{t-p}$$

$$\begin{bmatrix} 1 & & & & 0 \\ -\hat{a}_1 & 1 & & & \\ \vdots & \ddots & \ddots & & \\ -\hat{a}_P & & \ddots & \ddots & \\ & \ddots & & \ddots & \ddots \\ 0 & & -\hat{a}_P & \cdots & -\hat{a}_1 & 1 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_T \end{bmatrix} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_T \end{bmatrix}$$

- 予測誤差 $\{\epsilon_t\}_{1 \leq t \leq T}$ と予測係数 $\{\hat{a}_p\}_{1 \leq p \leq P}$ から元信号を復元可能

$$s_1 = \epsilon_1$$

$$s_2 = \epsilon_2 + \hat{a}_1 s_1$$

$$s_3 = \epsilon_3 + \hat{a}_1 s_2 + \hat{a}_2 s_1$$

$$\vdots$$

(補足)相関係数算出における日米の差1

■ 予測残差の算出を厳密に考えてみる

- 区間 $[n_0, n_1]$ における $\varepsilon(n)$ の2乗和 η

$$\begin{aligned}\eta &= \sum_{n=n_0}^{n_1} \varepsilon(n)^2 = \sum_{n=n_0}^{n_1} \left(\sum_{i=0}^p \alpha_i x(n-i) \right)^2 \\ &= \sum_{n=n_0}^{n_1} \sum_{i=0}^p \sum_{j=0}^p \alpha_i \alpha_j x(n-i) x(n-j) \\ &= \sum_{i=0}^p \sum_{j=0}^p \alpha_i \alpha_j \nu_{ij}\end{aligned}$$

$$\text{where } \nu_{ij} = \sum_{n=n_0}^{n_1} x(n-i) x(n-j)$$

自己相関関数

(補足)相関係数算出における日米の差2

■予測残差の最小化

- 2乗残差和 η を最小にする α_j を求める

$$\frac{\partial \eta}{\partial \alpha_j} = 2 \sum_{i=0}^p \alpha_i v_{ij} = 0, \quad (j = 1, 2, \dots, p)$$

$$\therefore \sum_{i=1}^p \alpha_i v_{ij} = -v_{0j}, \quad (j = 1, 2, \dots, p) \quad \cdots(2)$$

よって、線形予測係数 α_i を算出するには、上記の p 個の連立1次方程式を解けばよい。

北米ATTベル研究所のAtalらはコレスキー分解で解いた

⇒ 必ずしも解が存在するとは限らない？

(補足)相関係数算出における日米の差3

■ 安定に解を求めるには...

- 自己相関関数 v_{ij} に制約を設ける

$$n_0 = -\infty, n_1 = +\infty$$

このとき $x(n) = 0$, if $(n < 0 \text{ or } N \leq n)$

$$v_{ij} = \sum_{n=-\infty}^{\infty} x(n)x(n+|i-j|) = r_{|i-j|}$$

i, j の2変数に関する関数が1変数 $|i-j|$ のみの関数となる。(板倉ら日本側のアプローチ)

- ・この場合の相関行列はテプリッツ型の行列となる。
- ・レビンソン・ダービン(板倉)アルゴリズム等の高速解法で解ける。
- ・かつそこで得られた線形予測係数による自己回帰モデルは安定。

線形予測符号化(Linear Predictive Coding)

■ 時系列信号の可逆圧縮符号化の標準的な方式



線形予測分析

■ 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

線形予測分析

■ 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

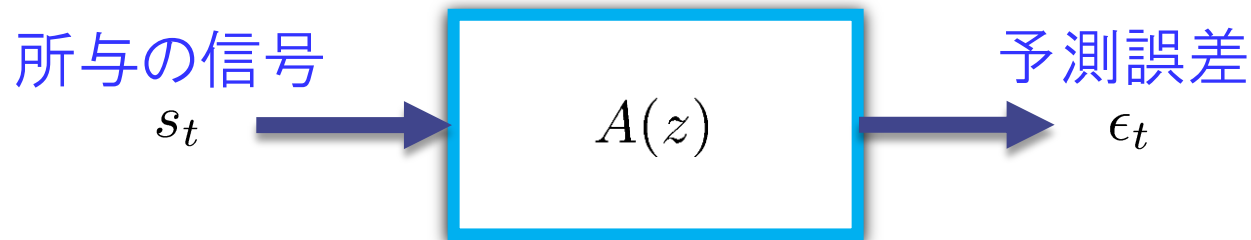
■ 最尤スペクトル推定

■ 板倉斎藤距離

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

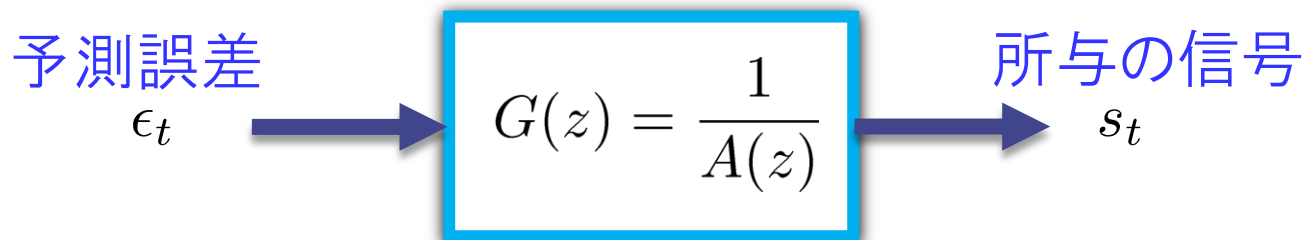
$$\epsilon_t = s_t - \sum_{p=1}^P a_p s_{t-p}$$



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

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)

音声生成過程のモデルとして

復習

人間頭部の断面図

②

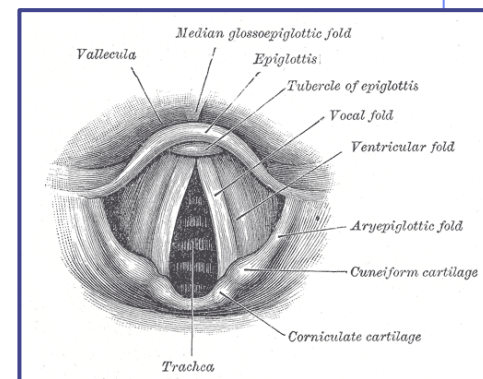
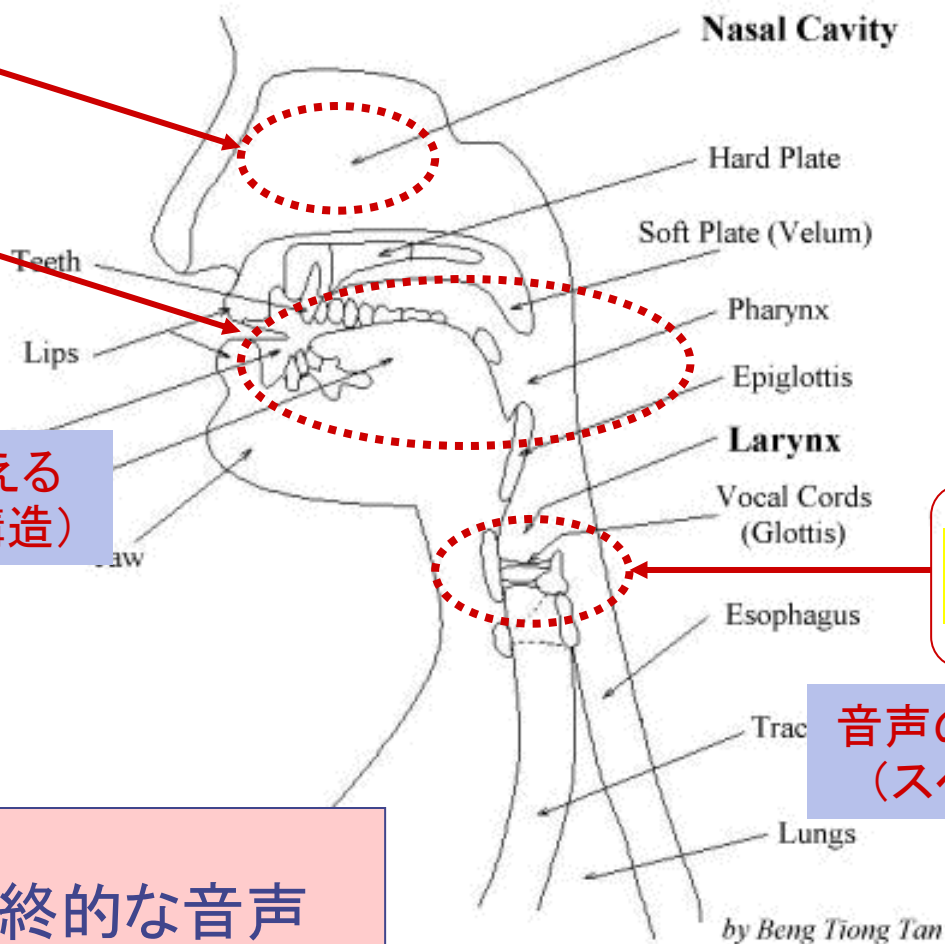
鼻腔

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

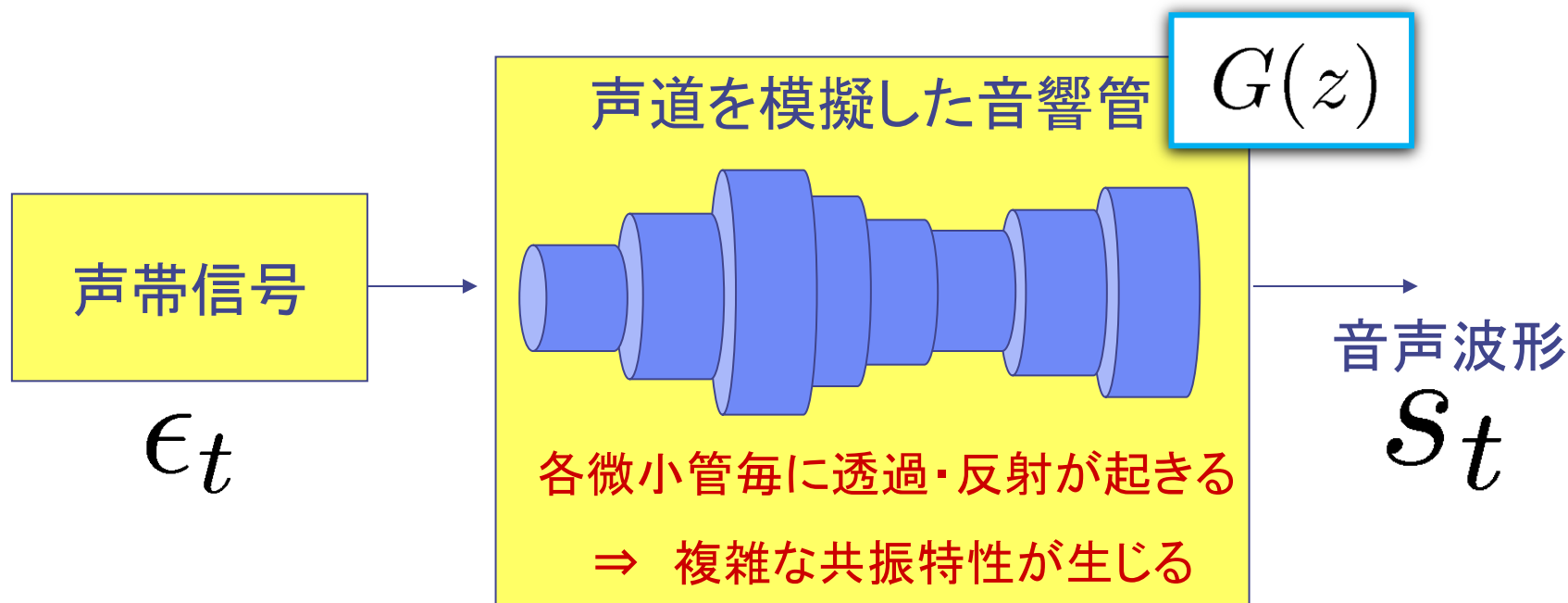
声帯の振動

音声の基本周期を与える
(スペクトル微細構造)

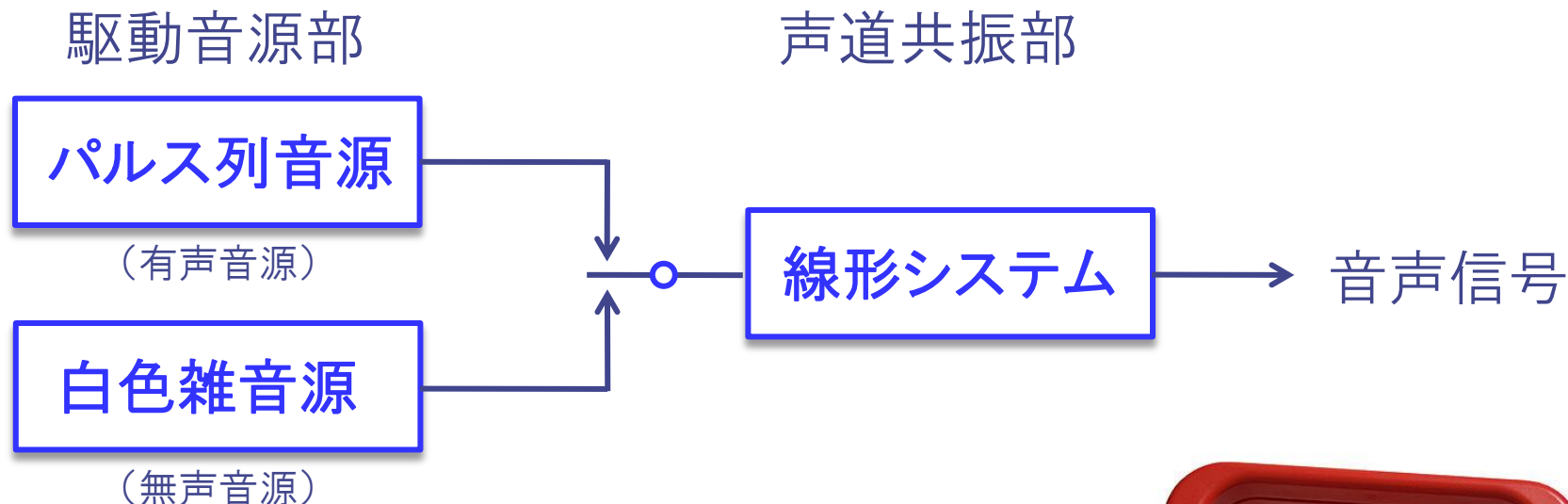
音声生成に適したモデルとは？

■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



音声生成の線形モデル



■ “Speak & Spell”

- LPCに基づく音声合成LSIを搭載
- 米国のTexas Instruments社開発
- 1978年発売

※因みに私(猿渡)はリアルタイム
(中学1年)で買って持っていました



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)

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

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 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

線形予測分析

■ 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

線形予測分析

■ 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

周波数領域での定式化

■ 時間領域では・・・

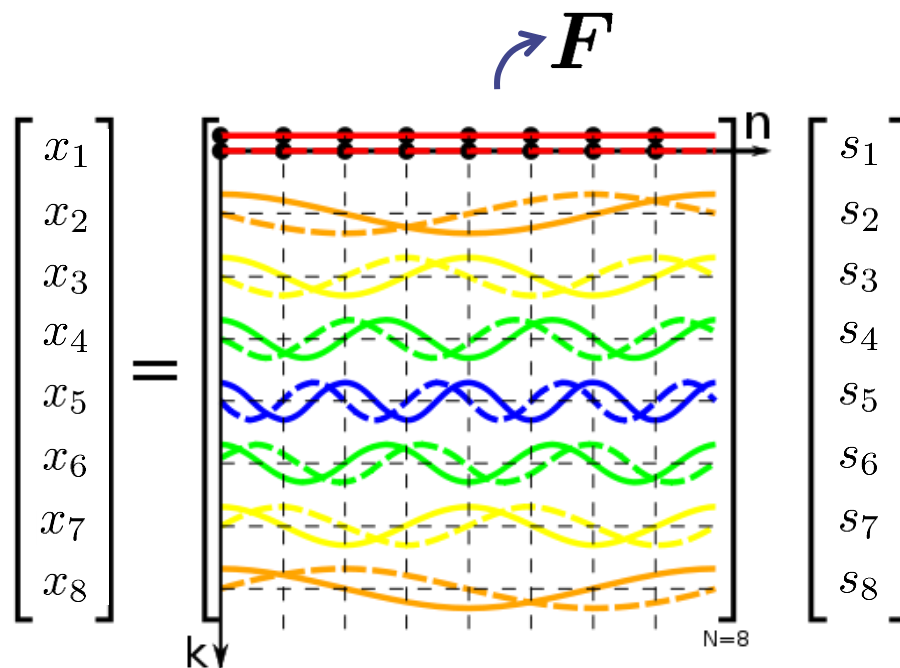
$$\left. \begin{array}{l} \blacksquare \Psi^{-1} \epsilon = s \\ \blacksquare \epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}) \end{array} \right\} \longrightarrow s \sim \mathcal{N}(\mathbf{0}, \sigma^2 \Psi^{-1} (\Psi^{-1})^T)$$

■ 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

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

■ 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...

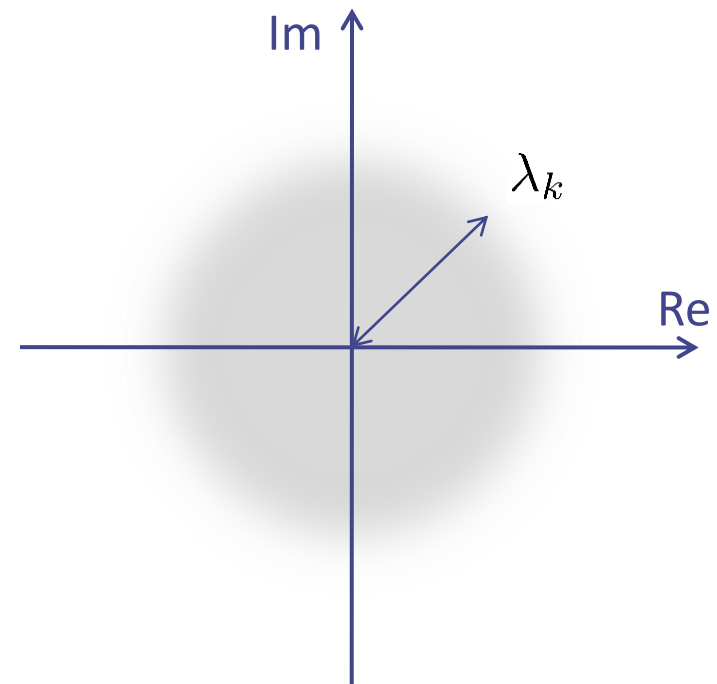
$$\mathbf{x} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \text{diag}(\lambda_1, \dots, \lambda_T))$$

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

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

- 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

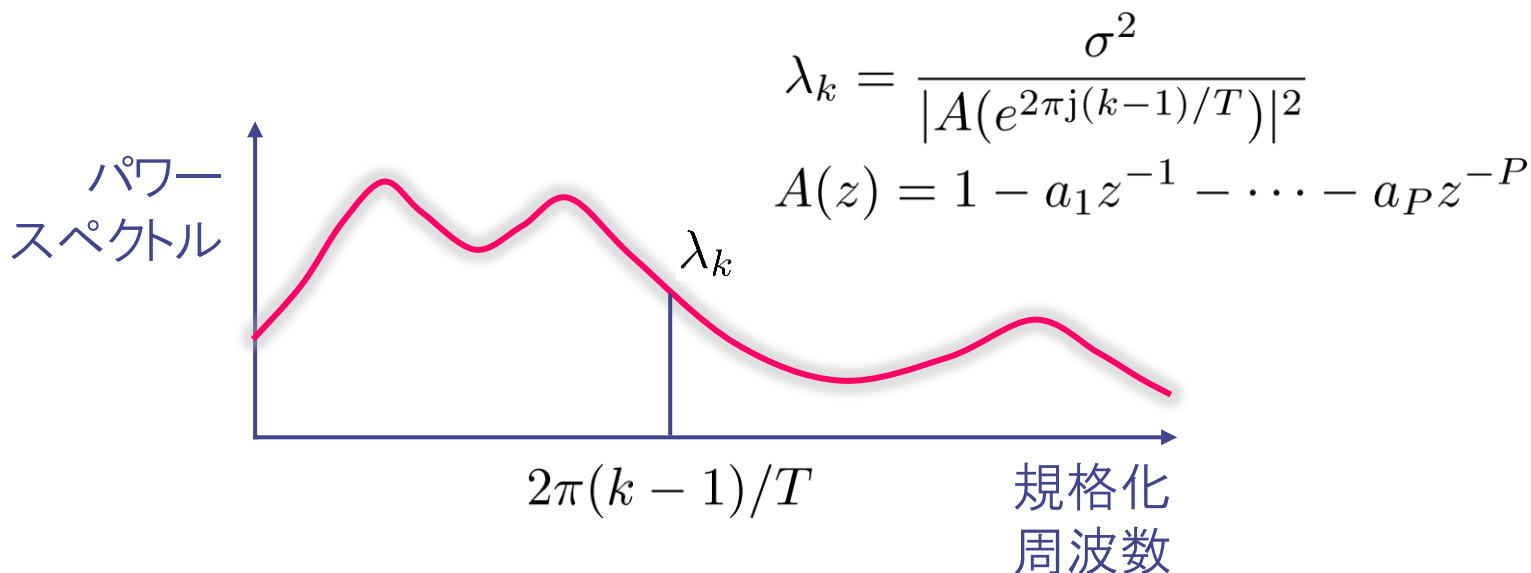
$$x_k \sim \mathcal{N}_{\mathbb{C}}(0, \lambda_k)$$



「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

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



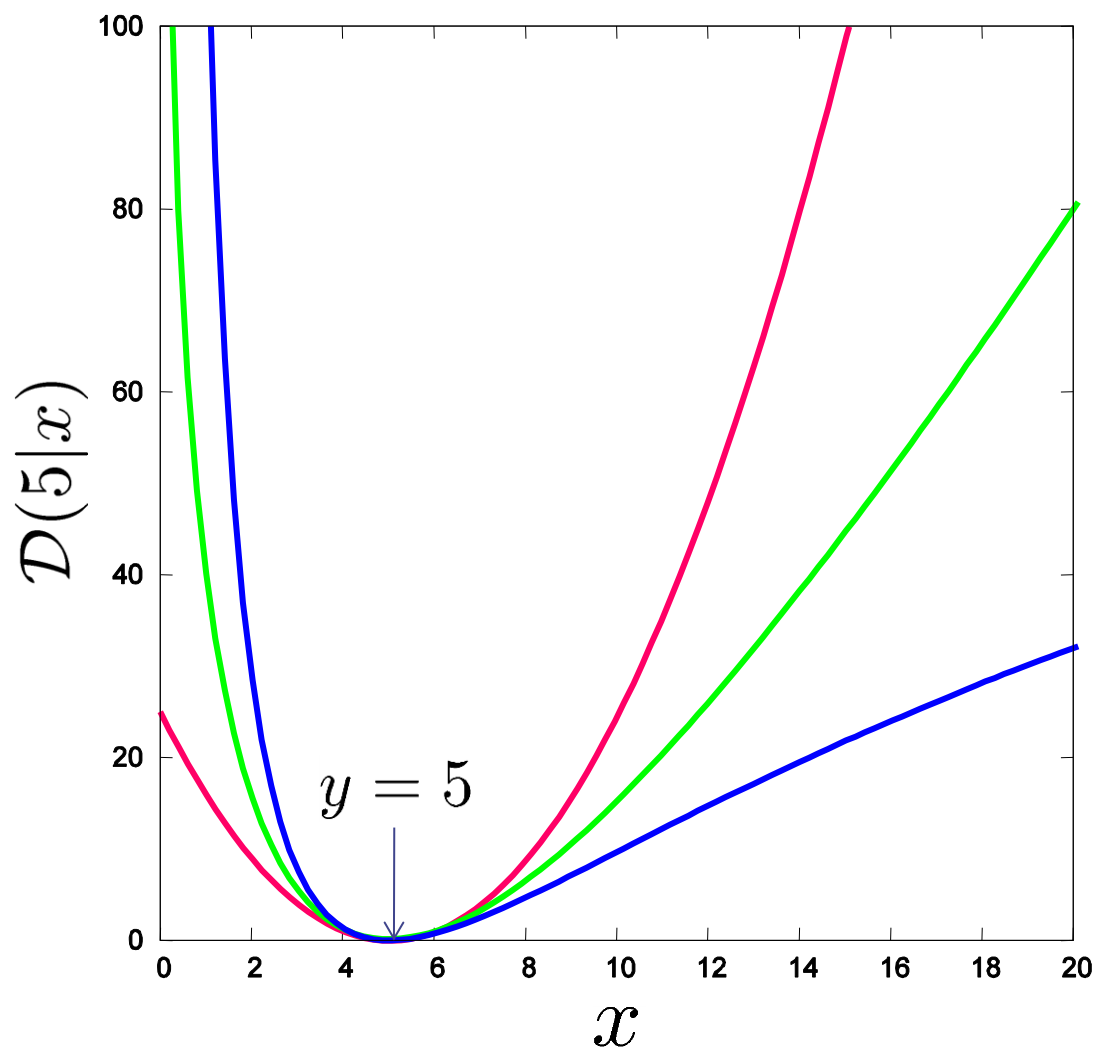
- 上記の対数尤度は、定数項を除けば以下と等しい

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

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

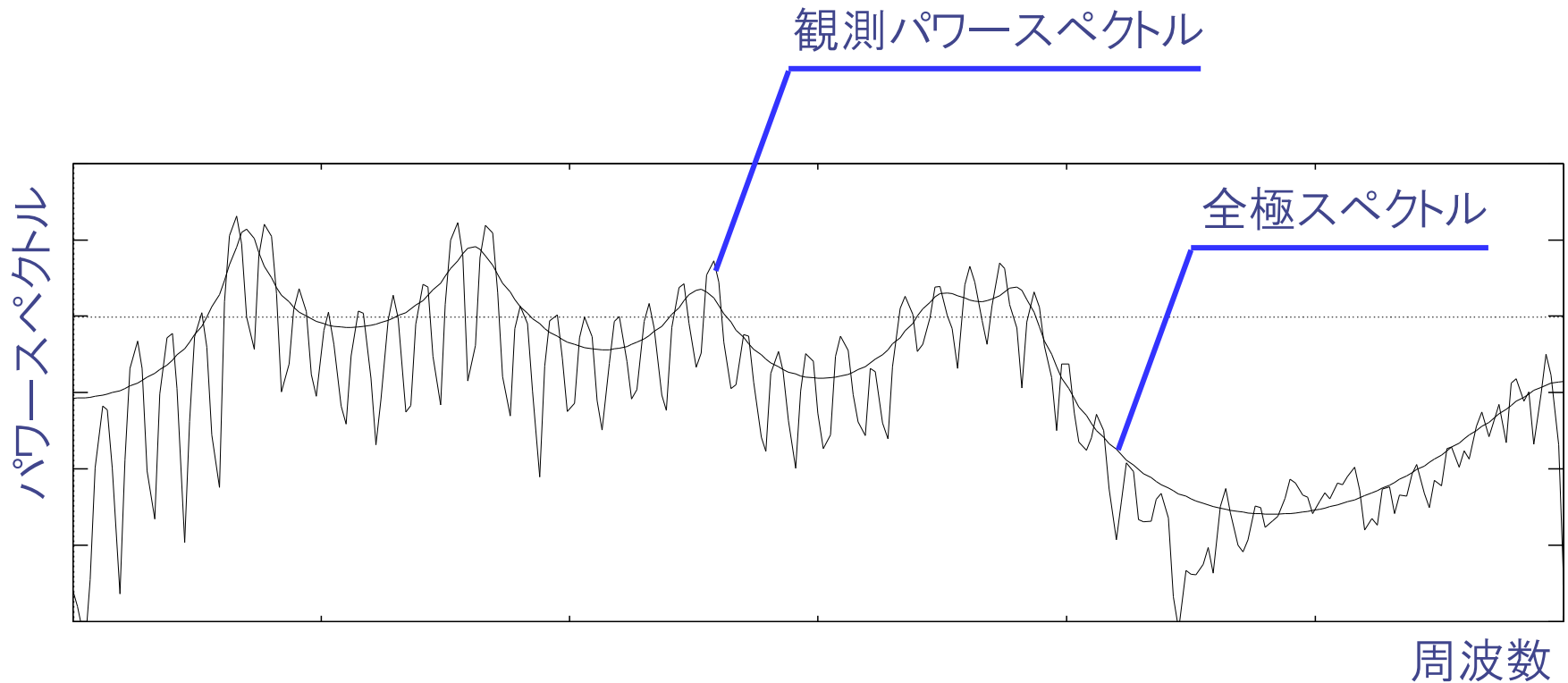
$$\mathcal{D}(x|y) = y \log \frac{y}{x} - (y - x)$$

— 板倉斎藤距離

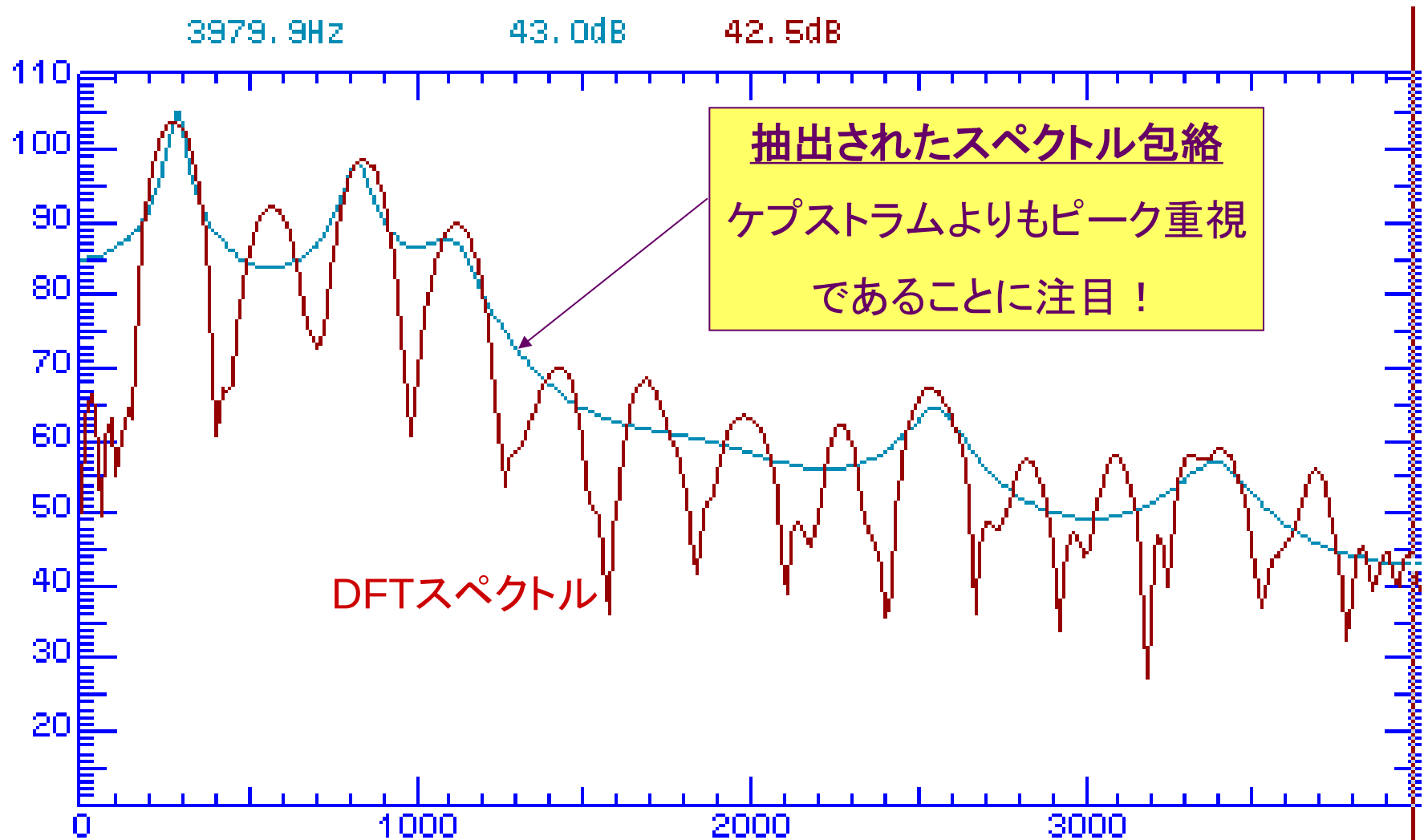
$$\mathcal{D}(x|y) = \frac{y}{x} - \log \frac{y}{x} - 1$$

線形予測分析は「スペクトル包絡」の推定に相当

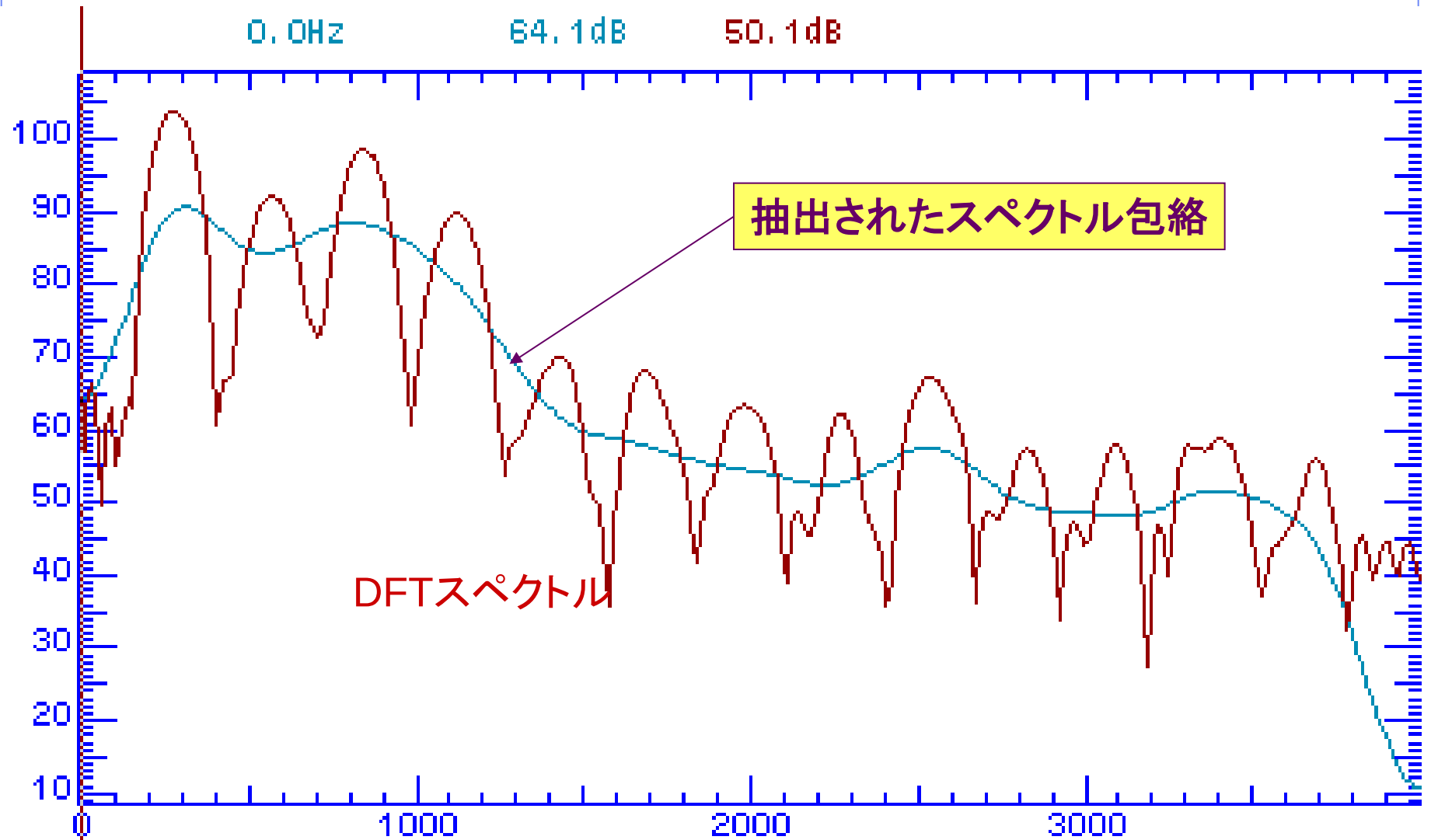
■ 観測パワースペクトルと全極スペクトルとの板倉斎藤距離最小化



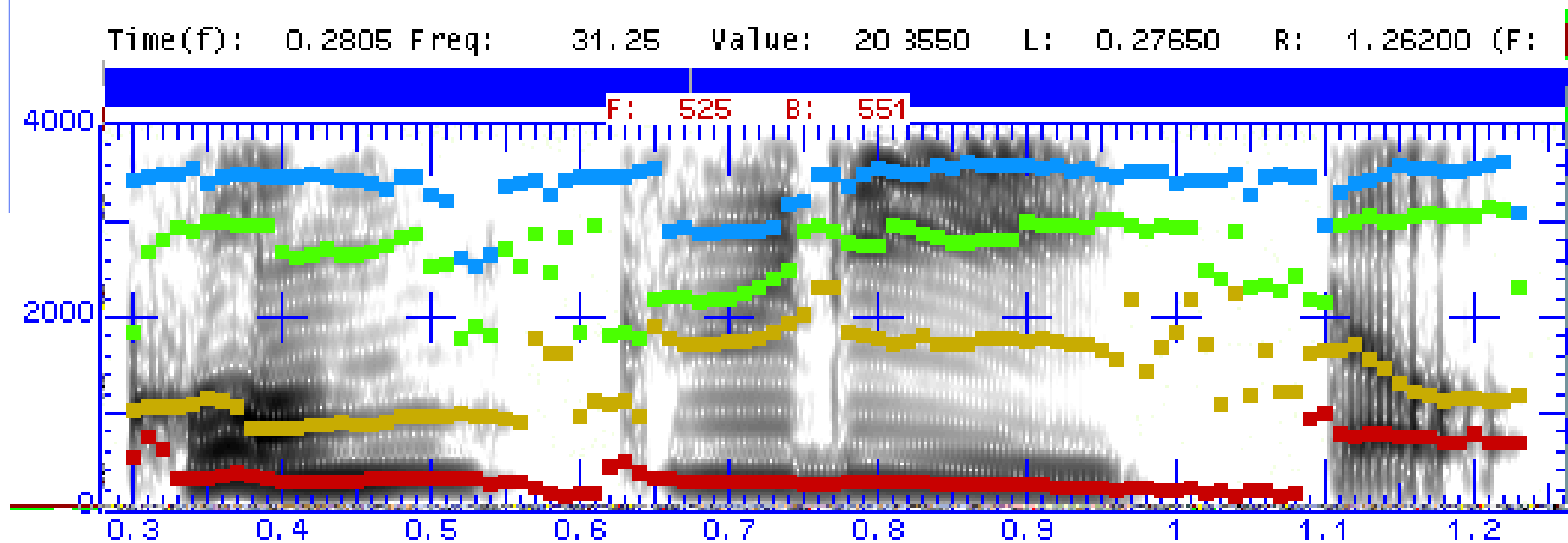
線形予測によるスペクトル包絡の良さは？



(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

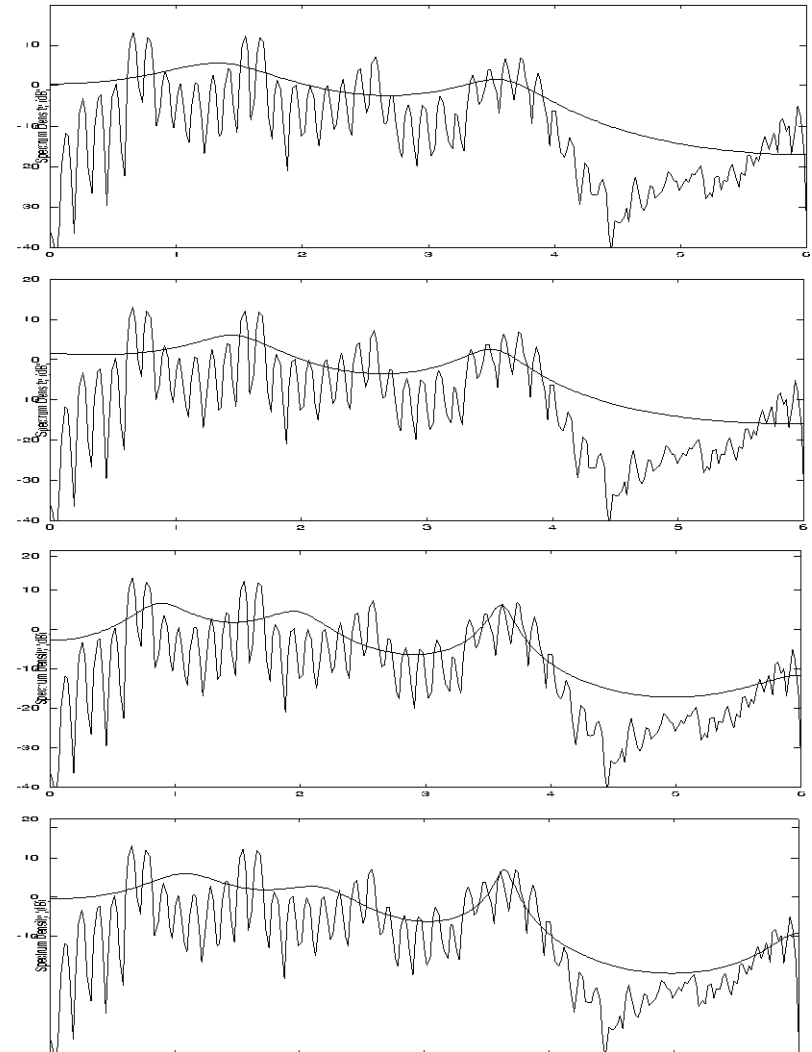
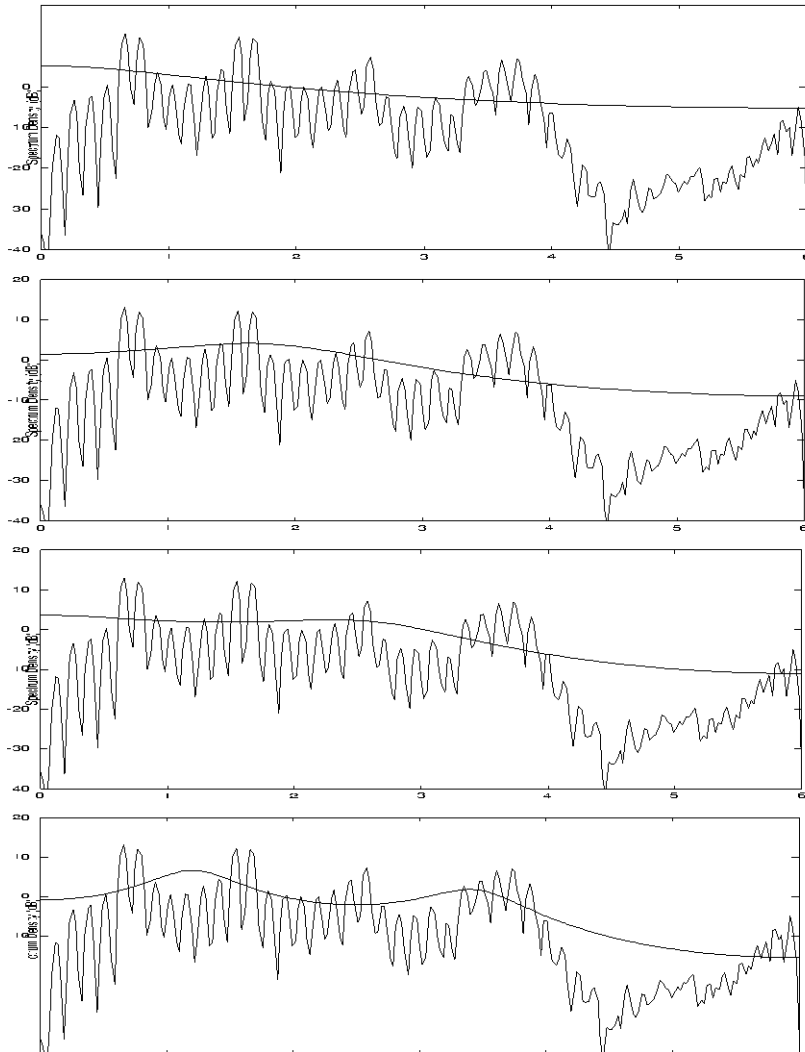
■日経産業新聞 1999年4月20日掲載

「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」**

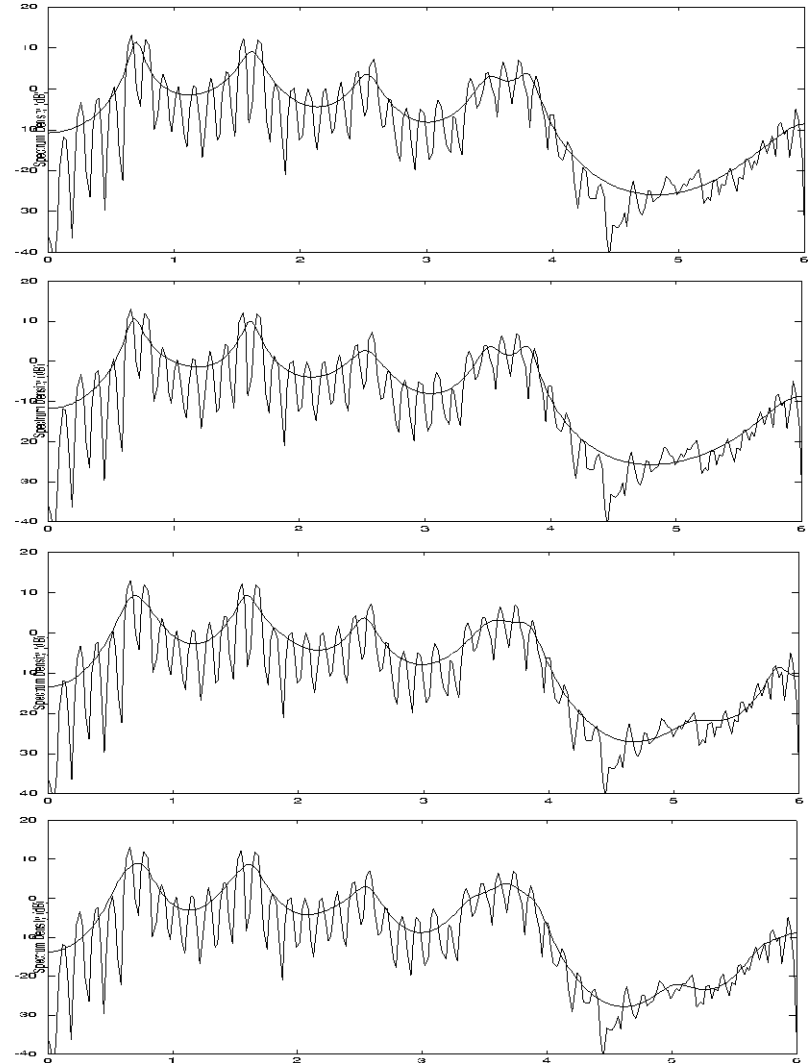
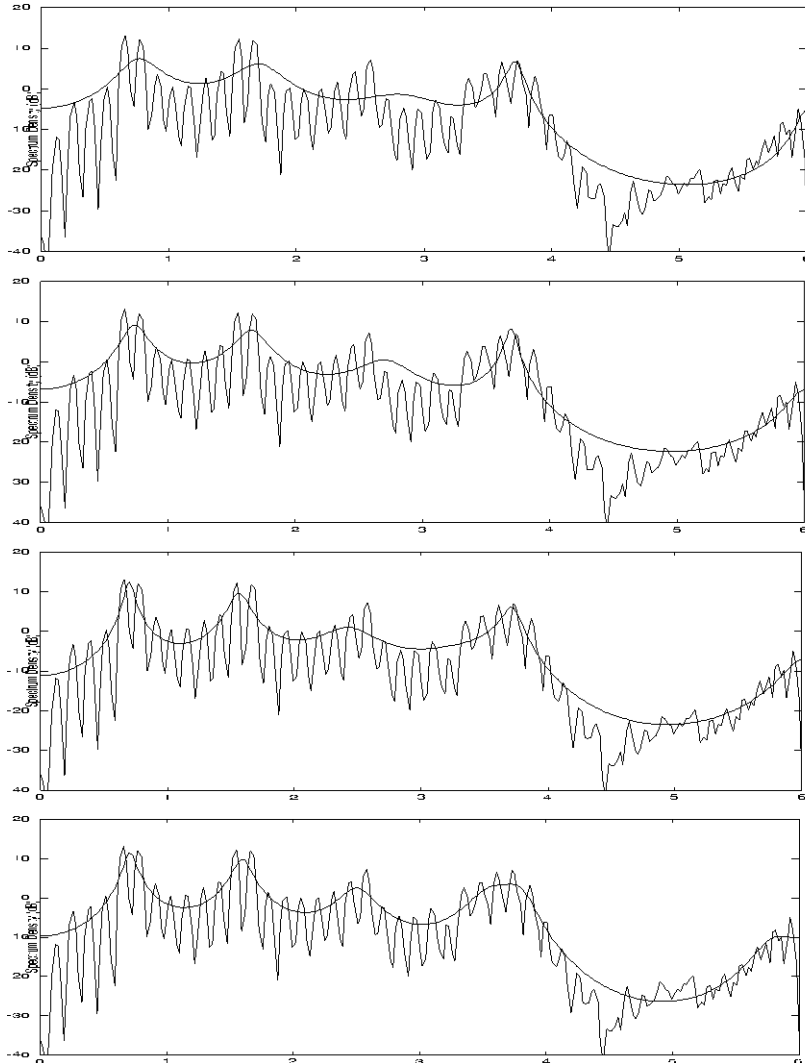
LPCによる音声スペクトル推定の例

■ 1～8次



LPCによる音声スペクトル推定の例

■ 9, 10, 12, 14, 16, 18, 22, 26次



線形予測分析のまとめ

■長所

- 高速解法が存在するため比較的単純な操作でスペクトル包絡抽出可能
- 抽出されたスペクトル包絡において、ホルマント共振がより強調される(cf. ケプストラム分析)
- より少ないパラメータ(たかだか p 個の予測係数のみ)で音声スペクトル包絡を表現可能 ⇒ 音声符号化に有利

■問題点

- 線形予測係数 α_i を量子化して伝送をする場合、**伝送誤差の影響によってすぐに不安定なフィルタになってしまう。**
(例) 典型的な電話音声の場合11 bits以上の精度必要
- 線形予測係数とスペクトルの直観的な関連がないので、スペクトルの補間を行う場合に**予測係数補間が不可能。**

線形予測分析の拡張1 (PARCOR)

■量子化誤差対策: PARCOR分析

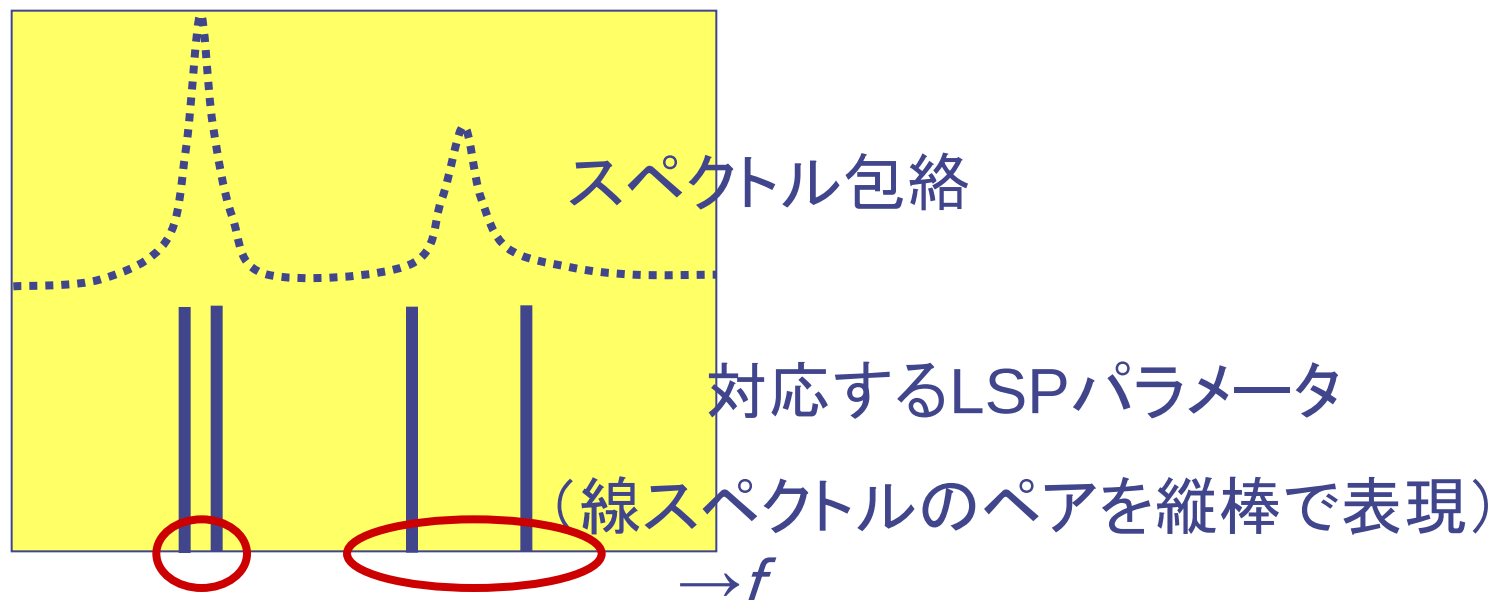
- 線形予測による伝達関数⇒音響管の共振モデルに対応
- 線形予測係数を音響管の各管における反射係数へ一意に変換可能
- 反射係数が1を超えることは無い⇒伝送エラーなどで歪んでしまった(1以上にバケてしまった)反射係数を近似回復できる。
つまり絶対安定な伝達関数を受信側で構成可能

■しかしまだ改善点が...

- より情報圧縮を行いたい場合、とびとびの時間分析フレームのデータのみを伝送し、受手側では時間補間をすることによって復元を行いたい。しかし、LPC係数・PARCOR係数とも、時間軸方向の連続性はあまり明確ではない。

線形予測分析の拡張2(LSP)

- 係数の時間補間対策: LSP(線スペクトル対)係数
 - PARCOR係数をさらに周波数領域へマッピング
 - ⇒ 絶対安定性を保ちつつ **スペクトルの時間補間が可能**



強い共振ピーク付近に棒線が密集。共振の強さは密集度合で決まる。

伝送するのは**線スペクトル(ペア)の周波数位置のみ**。

LSP係数による時間補間

