

# 信号処理論特論

## 第9回 (12/6)

情報理工学系研究科 猿渡 洋  
hiroshi\_saruwatari@ipc.i.u-tokyo.ac.jp

# 講義予定

- 9/27: 第1回 統計的音声音響信号処理概論
- 10/04: 第2回 非負値行列因子分解
- 10/11: 第3回 独立因子分析(ICA, IVA, ILRMA)
- 10/18: 第4回 独立因子分析(続き)
- 10/25: 第5回 音場再現・スパース最適化
- 11/01: 第6回 音声合成・変換1
- 11/15: 第7回 【レポート課題1】
- 11/22: 第8回 話者認識
- 11/29: 休講
- 12/06: 第9回 エンハンスメント・高次統計量解析
- 12/13: 休講
- 12/20: 第10回 音声合成・変換2
- 01/10: 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なってます)

## ■ 成績評価

- 出席点
- レポート点(2回の提出が必須)

# 本日の話題

## ■前半：音声強調信号処理の概観

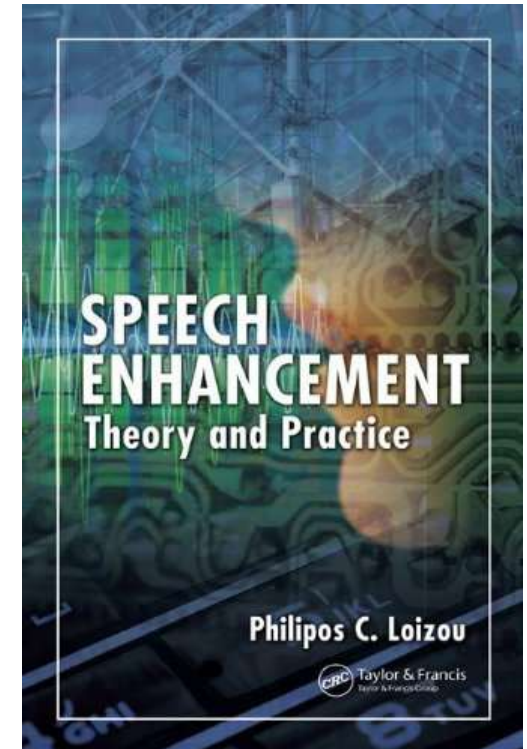
- 雑音や残響が重畳された観測信号から統計的に音声信号を強調したり抽出したりする技術
- 音声アプリケーション(携帯電話, 音声認識, 航空通信, テレ会議システム, 補聴器利用者による音声通信)において特に有用

## ■後半：聴感的品質制御について

- 音声強調処理後の「聴感品質」を制御
- 高次統計量と聴感品質の関係
- 音声強調処理系を他の信号処理と組み合わせた時の最適化問題

## 【参考文献】

- P. C. Loizou, *Speech Enhancement: Theory and practice*, CRC Press, 2007.



# 前半の目次

- 音声強調問題の設定
  - 解法の分類
    - 推定対象
      - ◆ 音声の複素スペクトル／振幅スペクトル
    - 推定量
      - ◆ 最尤(ML)推定量／最小平均二乗誤差(MMSE)推定量
- 解法1: 複素スペクトルのMMSE推定量 [Wiener1949]
  - 解法2: 振幅スペクトルのML推定量 [McAulay1980]
  - 解法3: 振幅スペクトルのMMSE推定量 [Ephraim1984]
- 事前SN比(または音声パワースペクトル密度)の推定方法
    - 最尤法(パワー減算(PS)法)

# 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

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

から音声に関するパラメータ  $\theta$  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  $\mathcal{N}(0, \lambda_X(\omega))$  に従う
- 雑音  $N(\omega)$  は平均0の複素正規分布  $\mathcal{N}(0, \lambda_N(\omega))$  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)

# 解法の種類

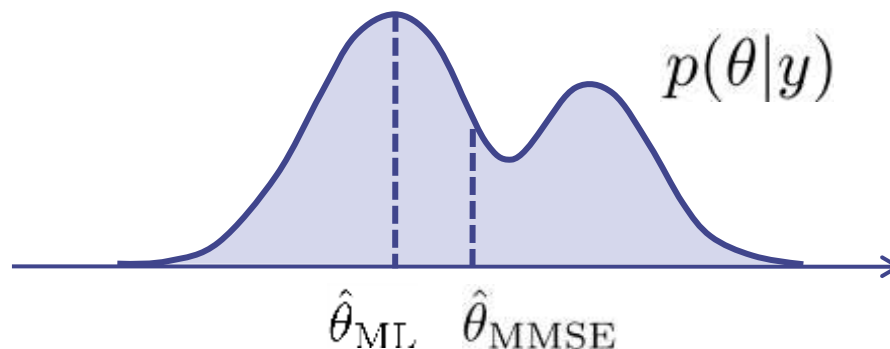
## ■ $\theta$ の推定量

- 最尤(Maximum Likelihood)推定量

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

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

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



## ■ 何を $\theta$ とするか？

- 音声スペクトル  $X(\omega)$
- 音声スペクトル  $X(\omega)$  の振幅  $A(\omega) := |X(\omega)|$

# 解法の分類

推定量 \ $\theta$	$X(\omega)$	$A(\omega) :=  X(\omega) $
ML	—	[R.J. McAulay and M.L. Malpass (1980)]
MMSE	[N. Wiener (1949)]	[Y. Ephraim and D. Malah (1984)]

N. Wiener (1949), Extrapolation, interpolation and smoothing of stationary time series with engineering applications, Cambridge, MA: MIT Press.

R.J. McAulay and M.L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," IEEE Trans. on Acoust. Speech Signal Process., Vol. 28, No. 2, pp. 137-145, 1980.

Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," IEEE Trans. Acoust. Speech Signal Process., Vol. 32, No. 6, pp. 1109-1121, 1984.



# 解法の分類

推定量 \ $\theta$	$X(\omega)$	$A(\omega) :=  X(\omega) $
ML	—	[R.J. McAulay and M.L. Malpass (1980)]
MMSE	[N. Wiener (1949)]	[Y. Ephraim and D. Malah (1984)]

N. Wiener (1949), *Extrapolation, interpolation and smoothing of stationary time series with engineering applications*, Cambridge, MA: MIT Press.

R.J. McAulay and M.L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Trans. on Acoust. Speech Signal Process.*, Vol. 28, No. 2, pp. 137-145, 1980.

Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust. Speech Signal Process.*, Vol. 32, No. 6, pp. 1109-1121, 1984.

# $X(\omega)$ のMMSE推定量 (1/2)

## ■ 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

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

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

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

## $X(\omega)$ のMMSE推定量 (2/2)

- Wienerフィルタは  $X(\omega)$  のMMSE推定量

$$\hat{X}_{\text{MMSE}}(\omega) = \underline{G_{\text{Wiener}}(\omega)} Y(\omega)$$

$$\begin{aligned} G_{\text{Wiener}}(\omega) &= \frac{\lambda_X(\omega)}{\lambda_X(\omega) + \lambda_N(\omega)} \\ &= \frac{\xi(\omega)}{\xi(\omega) + 1} \end{aligned}$$

$$\xi(\omega) = \frac{\lambda_X(\omega)}{\lambda_N(\omega)}$$

事前SN比

# 解法の分類

推定量 \ $\theta$	$X(\omega)$	$A(\omega) :=  X(\omega) $
ML	—	[R.J. McAulay and M.L. Malpass (1980)]
MMSE	[N. Wiener (1949)]	[Y. Ephraim and D. Malah (1984)]

N. Wiener (1949), Extrapolation, interpolation and smoothing of stationary time series with engineering applications, Cambridge, MA: MIT Press.

R.J. McAulay and M.L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," IEEE Trans. on Acoust. Speech Signal Process., Vol. 28, No. 2, pp. 137-145, 1980.

Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," IEEE Trans. Acoust. Speech Signal Process., Vol. 32, No. 6, pp. 1109-1121, 1984.

# $A(\omega)$ のML推定量 (1/4)

## ■ 問題設定:

- $Y(\omega) = A(\omega)e^{j\phi(\omega)} + N(\omega)$
- $N(\omega) \sim \mathcal{N}_{\mathbb{C}}(N(\omega); 0, \lambda_N(\omega)), \quad \phi(\omega) \sim \text{Uniform}[0, 2\pi)$

## ■ 求めたいのは $\hat{A}_{\text{ML}}(\omega) = \underset{A(\omega)}{\text{argmax}} \underline{p(Y(\omega)|A(\omega))}$

$$\underline{p(Y(\omega)|A(\omega), \phi(\omega))}$$

$$= \mathcal{N}_{\mathbb{C}}(Y(\omega); A(\omega)e^{j\phi(\omega)}, \lambda_N(\omega))$$

$$= \frac{1}{\pi\lambda_N(\omega)} \exp\left(-\frac{|Y(\omega) - A(\omega)e^{j\phi(\omega)}|^2}{\lambda_N(\omega)}\right)$$

$$= \frac{1}{\pi\lambda_N(\omega)} \exp\left(-\frac{|Y(\omega)|^2 - 2A(\omega)\text{Re}\{e^{-j\phi(\omega)}Y(\omega)\} + A(\omega)^2}{\lambda_N(\omega)}\right)$$

$$\underline{p(Y(\omega)|A(\omega))} = \frac{1}{2\pi} \int_0^{2\pi} \underline{p(Y(\omega)|A(\omega), \phi(\omega))} d\phi(\omega)$$

## $A(\omega)$ のML推定量 (2/4)

■  $\phi(\omega)$  の周辺化:

$$\begin{aligned} & \frac{1}{2\pi} \int p(Y(\omega) | A(\omega), \phi(\omega)) d\phi(\omega) \\ &= \frac{1}{\pi \lambda_N(\omega)} \int_0^{2\pi} \exp\left(-\frac{|Y(\omega)|^2 - 2A(\omega)\text{Re}\{e^{-j\phi(\omega)}Y(\omega)\} + A(\omega)^2}{\lambda_N(\omega)}\right) d\phi(\omega) \\ &= \frac{1}{\pi \lambda_N(\omega)} \exp\left(-\frac{|Y(\omega)|^2 + A(\omega)^2}{\lambda_N(\omega)}\right) \\ & \quad \cdot \frac{1}{2\pi} \int_0^{2\pi} \exp\left(\frac{2A(\omega)\text{Re}\{e^{-j\phi(\omega)}Y(\omega)\}}{\lambda_N(\omega)}\right) d\phi(\omega) \end{aligned}$$

第1種変形Bessel関数:

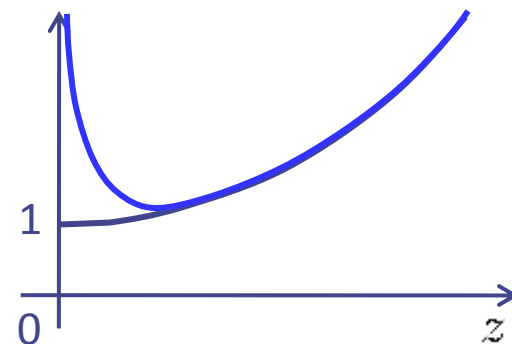
$$I_0(|z|) = \frac{1}{2\pi} \int_0^{2\pi} \exp(\text{Re}\{ze^{-j\phi}\}) d\phi$$

$$I_0\left(\frac{2A(\omega)|Y(\omega)|}{\lambda_N(\omega)}\right)$$

# $A(\omega)$ のML推定量 (3/4)

第一種変形Bessel関数の近似:

$$I_0(|z|) \simeq \frac{1}{\sqrt{2\pi|z|}} \exp(|z|)$$



$$\underline{I_0\left(\frac{2A(\omega)|Y(\omega)|}{\lambda_N(\omega)}\right)} \simeq \frac{1}{\sqrt{2\pi \frac{2A(\omega)|Y(\omega)|}{\lambda_N(\omega)}}} \exp\left(\frac{2A(\omega)|Y(\omega)|}{\lambda_N(\omega)}\right)$$

■  $p(Y(\omega)|A(\omega))$

$$\simeq \frac{1}{\pi \lambda_N(\omega)} \frac{1}{\sqrt{2\pi \frac{2A(\omega)|Y(\omega)|}{\lambda_N(\omega)}}} \exp\left(-\frac{|Y(\omega)|^2 + A(\omega)^2 - 2A(\omega)|Y(\omega)|}{\lambda_N(\omega)}\right)$$

■ 求めたかったのは  $\hat{A}_{\text{ML}}(\omega) = \underset{A(\omega)}{\operatorname{argmax}} p(Y(\omega)|A(\omega))$

$$\frac{\partial \log p(Y(\omega)|A(\omega))}{\partial A(\omega)} = 0 \quad \rightarrow \quad \hat{A}_{\text{ML}}(\omega) = \frac{1}{2} \left( |Y(\omega)| + \sqrt{|Y(\omega)|^2 - \lambda_N(\omega)} \right)$$

## $A(\omega)$ のML推定量 (4/4)

- 最尤の  $A(\omega)$  と観測位相  $Y(\omega)/|Y(\omega)|$  により  $X(\omega)$  の推定値を得る

$$\hat{X}_{\text{McAulay}}(\omega) = \hat{A}_{\text{ML}}(\omega) \frac{Y(\omega)}{|Y(\omega)|} = \underline{G_{\text{McAulay}}(\omega)} Y(\omega)$$

$$\underline{G_{\text{McAulay}}(\omega)} = \frac{\hat{A}_{\text{ML}}(\omega)}{|Y(\omega)|}$$

$$= \frac{1}{2} + \frac{1}{2} \sqrt{\frac{|Y(\omega)|^2 - \lambda_N(\omega)}{|Y(\omega)|^2}}$$

$$= \frac{1}{2} + \frac{1}{2} \sqrt{\frac{\gamma(\omega) - 1}{\gamma(\omega)}}$$

$$\gamma(\omega) = \frac{|Y(\omega)|^2}{\lambda_N(\omega)}$$

事後SN比



# 解法の分類

推定量 \ $\theta$	$X(\omega)$	$A(\omega) :=  X(\omega) $
ML	—	[R.J. McAulay and M.L. Malpass (1980)]
MMSE	[N. Wiener (1949)]	[Y. Ephraim and D. Malah (1984)]

N. Wiener (1949), Extrapolation, interpolation and smoothing of stationary time series with engineering applications, Cambridge, MA: MIT Press.

R.J. McAulay and M.L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," IEEE Trans. on Acoust. Speech Signal Process., Vol. 28, No. 2, pp. 137-145, 1980.

Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," IEEE Trans. Acoust. Speech Signal Process., Vol. 32, No. 6, pp. 1109-1121, 1984.

# $A(\omega)$ のMMSE推定量 (1/4)

## ■ 問題設定:

- $Y(\omega) = X(\omega) + N(\omega)$ ,  $X(\omega) = A(\omega)e^{j\phi(\omega)}$
- $N(\omega) \sim \mathcal{N}_{\mathbb{C}}(N(\omega); 0, \lambda_N(\omega))$ ,  $X(\omega) \sim \mathcal{N}_{\mathbb{C}}(X(\omega); 0, \lambda_X(\omega))$

## ■ 求めたいのは $\hat{A}_{\text{MMSE}}(\omega) = \mathbb{E}[A(\omega)|Y(\omega)]$

$$\begin{aligned} p(A(\omega)|Y(\omega)) &= \frac{p(Y(\omega)|A(\omega))p(A(\omega))}{p(Y(\omega))} \\ &= \frac{p(Y(\omega)|A(\omega))p(A(\omega))}{\int_0^{\infty} p(Y(\omega)|A(\omega))p(A(\omega))dA(\omega)} \end{aligned}$$

$$\begin{aligned} \hat{A}_{\text{MMSE}}(\omega) &= \int_0^{\infty} A(\omega)p(A(\omega)|Y(\omega))dA(\omega) \\ &= \frac{\int_0^{\infty} A(\omega)\underline{p(Y(\omega)|A(\omega))p(A(\omega))}dA(\omega)}{\int_0^{\infty} \underline{p(Y(\omega)|A(\omega))p(A(\omega))}dA(\omega)} \end{aligned}$$

## $A(\omega)$ のMMSE推定量 (2/4)

$$\blacksquare \underline{p(Y(\omega)|A(\omega))p(A(\omega))} = \int_0^{2\pi} \underline{p(Y(\omega)|A(\omega), \phi(\omega))p(A(\omega), \phi(\omega))} d\phi$$

$$\underline{p(Y(\omega)|A(\omega), \phi(\omega))} = \mathcal{N}_{\mathbb{C}}(Y(\omega); A(\omega)e^{j\phi(\omega)}, \lambda_N(\omega))$$

$$\underline{p(A(\omega), \phi(\omega))} = \mathcal{N}_{\mathbb{C}}(A(\omega)e^{j\phi(\omega)}; 0, \lambda_X(\omega))$$

$$\hat{A}_{\text{MMSE}}(\omega) =$$

$$\frac{\int_0^{\infty} \int_0^{2\pi} A(\omega) \mathcal{N}_{\mathbb{C}}(Y(\omega); A(\omega)e^{j\phi(\omega)}, \lambda_N(\omega)) \mathcal{N}_{\mathbb{C}}(A(\omega)e^{j\phi(\omega)}; 0, \lambda_X(\omega)) d\phi(\omega) dA(\omega)}{\int_0^{\infty} \int_0^{2\pi} \mathcal{N}_{\mathbb{C}}(Y(\omega); A(\omega)e^{j\phi(\omega)}, \lambda_N(\omega)) \mathcal{N}_{\mathbb{C}}(A(\omega)e^{j\phi(\omega)}; 0, \lambda_X(\omega)) d\phi(\omega) dA(\omega)}$$
$$= \sqrt{\frac{\lambda_X(\omega)\lambda_N(\omega)}{\lambda_X(\omega) + \lambda_N(\omega)}} \Gamma(1.5) \Phi(-0.5, 1; \underline{\nu(\omega)})$$

$\Gamma$  : ガンマ関数

$\Phi$  : 合流超幾何関数

(confluent hypergeometric function)

$$\underline{\nu(\omega)} = \frac{\underline{\xi(\omega)}}{1 + \underline{\xi(\omega)}} \underline{\gamma(\omega)}$$

事前SN比      事後SN比

## $A(\omega)$ のMMSE推定量 (3/4)

$$\begin{aligned} \sqrt{\frac{\lambda_X(\omega)\lambda_N(\omega)}{\lambda_X(\omega) + \lambda_N(\omega)}} &= \sqrt{\frac{\lambda_X(\omega)}{1 + \xi(\omega)}} \\ &= \sqrt{\frac{\xi(\omega)|Y(\omega)|^2}{(1 + \xi(\omega))\gamma(\omega)} \frac{\gamma(\omega)}{\gamma(\omega)}} \\ &= \sqrt{\frac{\xi(\omega)\gamma(\omega)}{(1 + \xi(\omega))\gamma(\omega)^2} \frac{1}{\gamma(\omega)} Y(\omega)} \\ &= \frac{\sqrt{\nu(\omega)}}{\gamma(\omega)} Y(\omega) \end{aligned}$$

$$\xi(\omega) = \frac{\lambda_X(\omega)}{\lambda_N(\omega)}$$

$$\gamma(\omega) = \frac{|Y(\omega)|^2}{\lambda_N(\omega)}$$

$$\begin{aligned} \lambda_X(\omega) &= \xi(\omega)\lambda_N(\omega) \\ &= \xi(\omega) \frac{|Y(\omega)|^2}{\gamma(\omega)} \end{aligned}$$

$$\nu(\omega) = \frac{\xi(\omega)}{1 + \xi(\omega)} \gamma(\omega)$$

$$\Phi(-0.5, 1; -z) = e^{-z/2} [(1+z)I_0(z/2) + zI_1(z/2)]$$



$$\begin{aligned} \hat{A}_{\text{MMSE}}(\omega) &= \frac{\sqrt{\pi}}{2} \frac{\sqrt{\nu(\omega)}}{\gamma(\omega)} \exp\left(-\frac{\nu(\omega)}{2}\right) \\ &\quad \left[ (1 + \nu(\omega)) I_0\left(\frac{\nu(\omega)}{2}\right) + \nu(\omega) I_1\left(\frac{\nu(\omega)}{2}\right) \right] Y(\omega) \end{aligned}$$

## $A(\omega)$ のMMSE推定量 (4/4)

- $A(\omega)$  のMMSE推定量におけるスペクトルゲイン

$$\hat{A}_{\text{MMSE}}(\omega) = \underline{G_{\text{Ephraim}}(\omega)} Y(\omega)$$

$$\underline{G_{\text{Ephraim}}(\omega)} = \frac{\sqrt{\pi}}{2} \frac{\sqrt{\nu(\omega)}}{\gamma(\omega)} \exp\left(-\frac{\nu(\omega)}{2}\right) \left[ (1 + \nu(\omega)) I_0\left(\frac{\nu(\omega)}{2}\right) + \nu(\omega) I_1\left(\frac{\nu(\omega)}{2}\right) \right]$$

# 事前SN比と事後SN比

- どのスペクトルゲイン関数も  $\xi(\omega)$ ,  $\gamma(\omega)$  によって決まる

- 事前SN比:

$$\xi(\omega) = \frac{\lambda_X(\omega)}{\lambda_N(\omega)} \quad (\text{目的信号のパワースペクトル密度が必要})$$

- 事後SN比:

$$\gamma(\omega) = \frac{|Y(\omega)|^2}{\lambda_N(\omega)} \quad (\text{観測パワースペクトルのみでOK})$$

- 事前SN比が未知のときは??

→事前SN比を推定する必要あり

# 事前SN比の推定法

## ■ 最尤法 (Power Subtraction法)

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

- $N(\omega) \sim \mathcal{N}_{\mathbb{C}}(N(\omega); 0, \lambda_N(\omega)), \quad X(\omega) \sim \mathcal{N}_{\mathbb{C}}(X(\omega); 0, \lambda_X(\omega))$

のもとで 定数

$$\hat{\lambda}_X(\omega) = \operatorname{argmax}_{\lambda_X(\omega)} p(Y(\omega) | \lambda_X(\omega))$$

を音声パワースペクトル密度の推定値とする単純なやり方

### ■ 仮定より

$$p(Y(\omega) | \lambda_X(\omega)) = \mathcal{N}_{\mathbb{C}}(0, \lambda_X(\omega) + \lambda_N(\omega))$$

- $\frac{\partial \log p(Y(\omega) | \lambda_X(\omega))}{\partial \lambda_X(\omega)} = 0 \longrightarrow \hat{\lambda}_X(\omega) = |Y(\omega)|^2 - \lambda_N(\omega)$

- 事前SN比の推定値は  $\dots \hat{\xi}(\omega) = \frac{\hat{\lambda}_X(\omega)}{\lambda_N(\omega)} = \underline{\gamma(\omega)} - 1$   
事後SN比

# PS法による振幅スペクトル推定値を用いた $X(\omega)$ の推定

- PS法による推定振幅  $\sqrt{|Y(\omega)|^2 - \lambda_N(\omega)}$  と観測位相  $Y(\omega)/|Y(\omega)|$  により  $X(\omega)$  の推定値を得ることも可能

$$\hat{X}_{\text{PS}}(\omega) = \sqrt{|Y(\omega)|^2 - \lambda_N(\omega)} \frac{Y(\omega)}{|Y(\omega)|} = \underline{G_{\text{PS}}(\omega)} Y(\omega)$$

$$\begin{aligned} \underline{G_{\text{PS}}(\omega)} &= \frac{\sqrt{|Y(\omega)|^2 - \lambda_N(\omega)}}{|Y(\omega)|} \\ &= \sqrt{\frac{\gamma(\omega) - 1}{\gamma(\omega)}} \end{aligned}$$

$$\gamma(\omega) = \frac{|Y(\omega)|^2}{\lambda_N(\omega)}$$

事後SN比

- PS法による音声パワースペクトル密度推定値を代入したWienerフィルタと比較してみると・・・

$$G_{\text{Wiener}}(\omega) = \frac{\lambda_X(\omega)}{\lambda_X(\omega) + \lambda_N(\omega)} = \frac{\xi(\omega)}{\xi(\omega) + 1} = \frac{\nu(\omega) - 1}{\nu(\omega)}$$

$$\xi(\omega) = \nu(\omega) - 1$$



# まとめ

- 音声強調問題の設定
- 解法の分類
  - 推定対象
    - ◆ 音声の複素スペクトル／振幅スペクトル
  - 推定量
    - ◆ 最尤(ML)推定量／最小平均二乗誤差(MMSE)推定量

- 解法1: 複素スペクトルのMMSE推定量 [Wiener1949]
- 解法2: 振幅スペクトルのML推定量 [McAulay1980]
- 解法3: 振幅スペクトルのMMSE推定量 [Ephraim1984]

- 事前SN比(または音声パワースペクトル密度)の推定方法
  - 最尤法(パワー減算(PS)法)

後半：  
高次統計量に基づく聴感品質の制御

# ミュージカルノイズの発生と数量化

# 音声強調処理の問題点 - アーチファクトの発生 -

- ◆ 統計的音声強調処理は、非線形雑音抑圧部において、不快なアーチファクト(これは**ミュージカルノイズ**と呼ばれる)が発生する。これは「人が聴く」用途へ適用する際に、大きな問題となってしまう。



- ◆ 信号処理を改良することで多少のミュージカルノイズ軽減は可能である。また、統計推定手法ごとに「聴感的印象」は異なる。つまり、統計的推定には「音の個性」がある。我々は**芸術的観点**から統計的推定問題を眺める。
- ◆ 本来、ミュージカルノイズに関しては、数理解析がほとんどなされていないという現状がある。よって、
  - (1) まずミュージカルノイズの定量指標を定める必要がある。
  - (2) 次に、ミュージカルノイズを低減する信号処理を開発する。

# 統計推定における音色の差を聞いてみよう！

白色ノイズの場合

観測音



人ごみノイズの場合

観測音

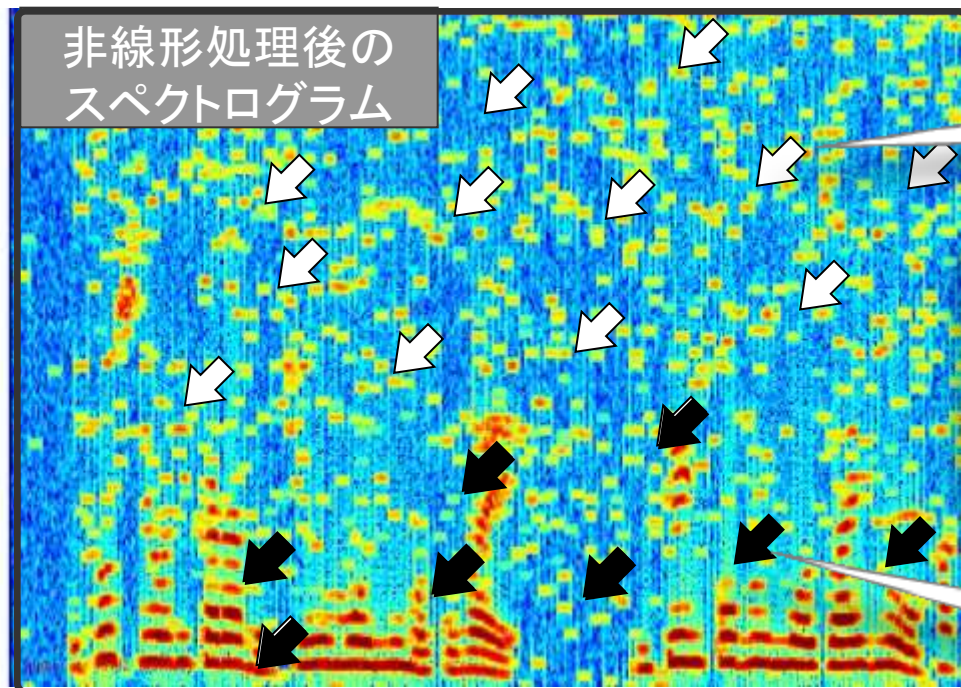


どの手法も推定エラー(残留雑音の分散)は同じです。

どの推定方式が「自然」だと感じましたか？

# ミュージカルノイズの原因成分

- ミュージカルノイズ = 「トーンを感じる歪み」
- トーンは、音声・音楽信号に豊富に含まれ、スペクトログラム上にて卓越成分があると、人は「トーンとして感じる」
- 音声・音楽信号以外で卓越した成分 = ミュージカルノイズ



雑音部分

ミュージカルノイズの評価  
= 卓越成分の評価

音声部分

# 高次統計量の導入

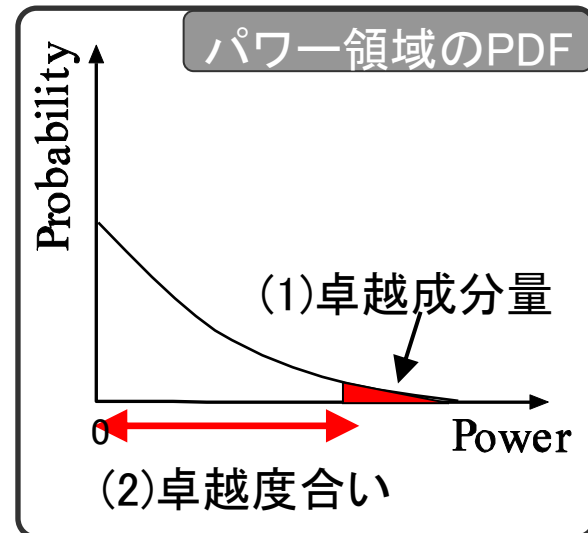
## ■ カートシス (尖度)

- 4次統計量に基づく「PDF裾野量尺度」
- 卓越成分の数値評価が可能となる

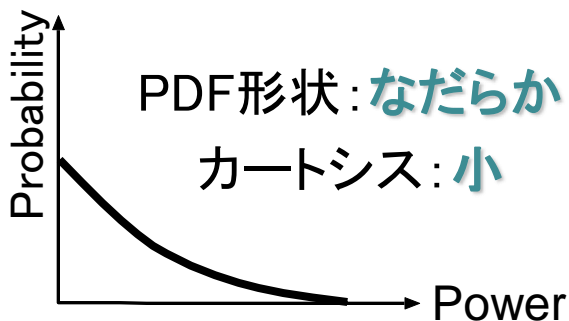
$$\text{kurtosis} = \frac{\mu_4}{\mu_2^2} \quad \mu_n \text{ はPDF } P(x) \text{ の } n \text{ 次モーメント}$$

$$\mu_n = \int_0^{\infty} x^n P(x) dx$$

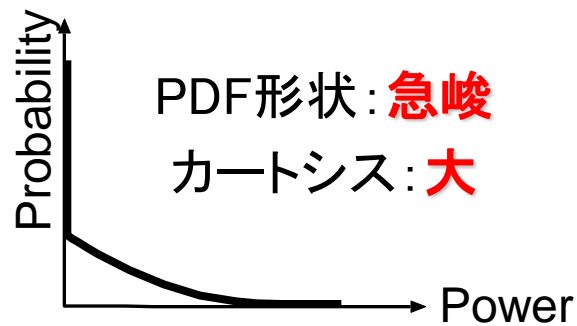
※ 中心化モーメントではないため厳密な尖度とは異なる



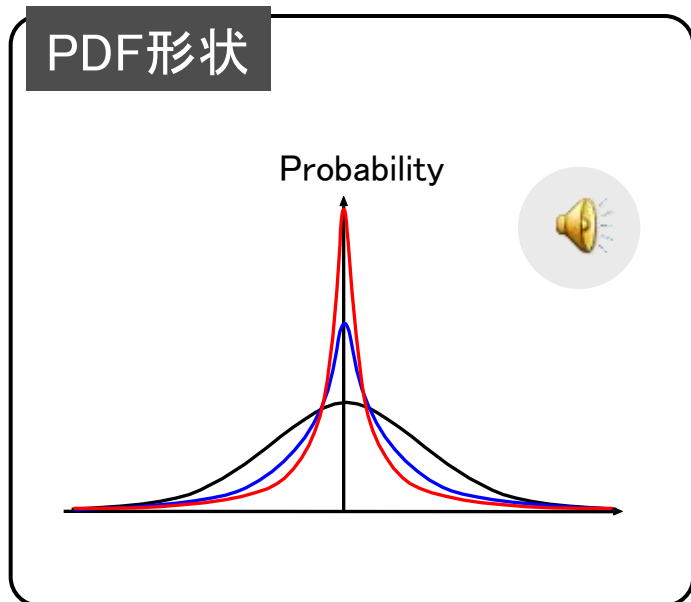
### PDF形状とカートシス



卓越成分が少ない



卓越成分が多い



ガウス分布



スーパーガウス分布（やや尖った形状）



スーパーガウス分布（より尖った形状）

と信号の従うPDF形状を変化させて  
作成した音源.

3秒ごとに、より急峻な分布に変わる.

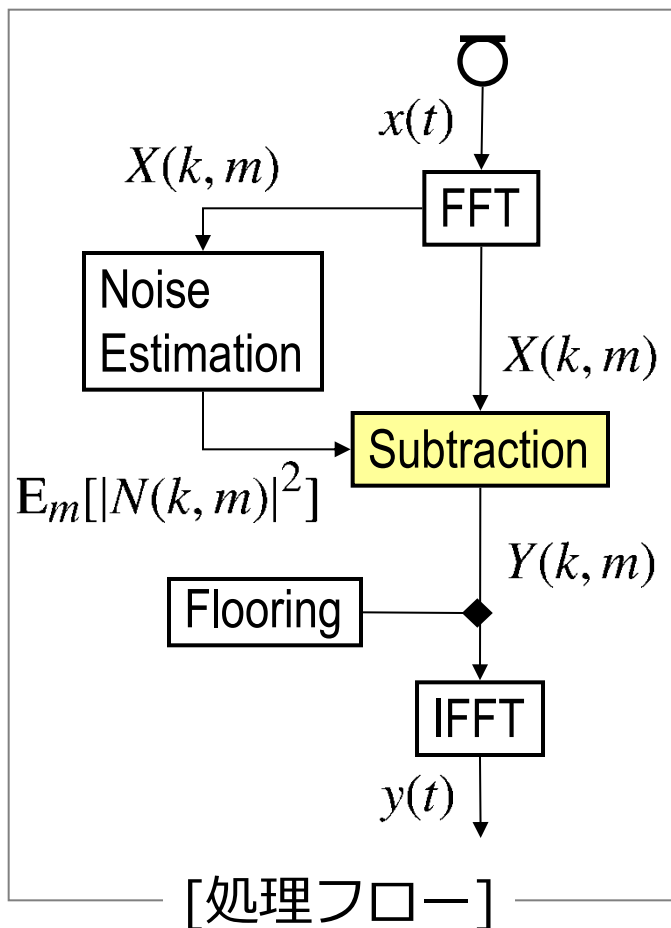
※ 非線形処理は一切行っていない

ミュージカルノイズはカートシスと強く関連していることが分かる



# スペクトル減算法 (Spectral Subtraction: SS)

## ■ パワードメインSS



パワードメインのSSでは,

$$Y(k, m) = \sqrt{|X(k, m)|^2 - \beta \cdot E_m[|N(k, m)|^2]} \cdot e^{j\arg(X(k, m))}$$

という式で出力を得る.

ここで, パワーが負のグリッドが生じた場合

$$\text{Flooring} \begin{cases} \text{if } |X(k, m)|^2 - \beta \cdot E_m[|N(k, m)|^2] < 0 \\ \text{then } Y(k, m) = \rho \cdot |X(k, m)| \cdot e^{j\arg(X(k, m))} \end{cases}$$

のように, フロアリングと呼ばれる処理で  
パワーを置き換えてやる.

パラメータは以下の通り.

$\beta$  : 減算係数 (処理強度パラメータ)

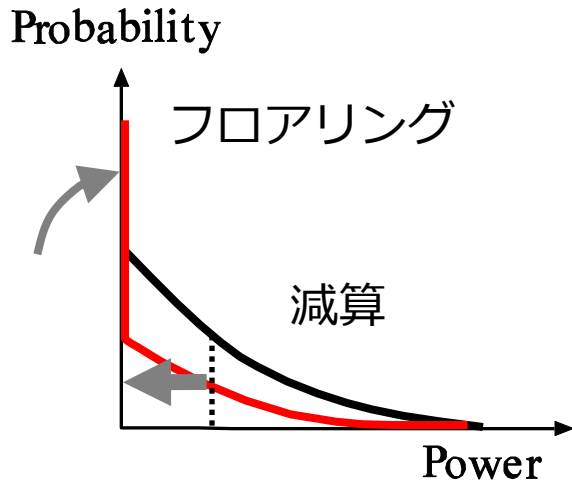
$\rho$  : フロアリング係数

# SSの運用とパラメータ

- パラメータはヒューリスティックに決定される
  - ⇒ 減算係数は通常, 1から2程度の値がよく用いられる
  - ⇒ フロアリング係数は通常, 0.1未満の値がよく用いられる
- SSの利用者は「SNR改善量」や「聞いた感じ」によって最適と思われる減算係数を決める
  - ⇒ SSの利用者は最適な減算係数を探した経験から,  
「音源ごとに最適と思われる減算係数が異なる」事を知っている
- フロアリングは強力過ぎた処理を和らげる働きをする
  - ⇒ 過減算したグリッドを観測信号の定数倍( $\rho$ )のパワーで補正する
  - ⇒ SNR改善量に対してマイナスに働く

# SSとPDFの変形

- トーン成分を統計的に評価するためにSSによる統計量の変化を明らかにする



$$\beta \cdot \mathbf{E}_m[|N(k, m)|^2]$$

処理後の信号のPDF

## Step 1

[SS処理] 雑音パワースペクトルの期待値を推定雑音パワースペクトルとする

## Step 2

[SS処理] 推定雑音を減算する  
[PDF変形] ゼロ方向へ平行移動する

## Step 3

[SS処理] フロアリング (パワーをゼロにおく)  
[PDF変形] パワーゼロ未満の確率をパワーゼロに積み重ねる

## ■ ガンマ分布

パワードメインの音声・雑音信号のモデリングによく使われる分布

$\alpha$  : 形状母数(shape parameter)

$\theta$  : 尺度母数(scale parameter) とすると,

ガンマ分布は,

$$P(x) = \frac{1}{\Gamma(\alpha) \theta^\alpha} \cdot x^{\alpha-1} e^{-\frac{x}{\theta}},$$

と表現される.

ただし、 $\begin{cases} x \geq 0, \alpha > 0, \theta > 0, \\ \Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx = (\alpha-1)\Gamma(\alpha-1) = (\alpha-1) \cdots (\alpha-j)\Gamma(\alpha-j). \end{cases}$

また, ガンマ分布の期待値は

$$E[P(x)] = \alpha\theta,$$

## ■ ガンマ分布の特徴

⇒ 片側分布のモデリングに適している

⇒  $\chi^2$ 分布などをモデリングできる

→ ガウス性雑音のパワースペクトルは $\chi^2$ 分布, 音声のパワースペクトルはより鋭い分布に従うことが知られている

⇒ ガンマ関数に基づいており数学的な利便性が高い

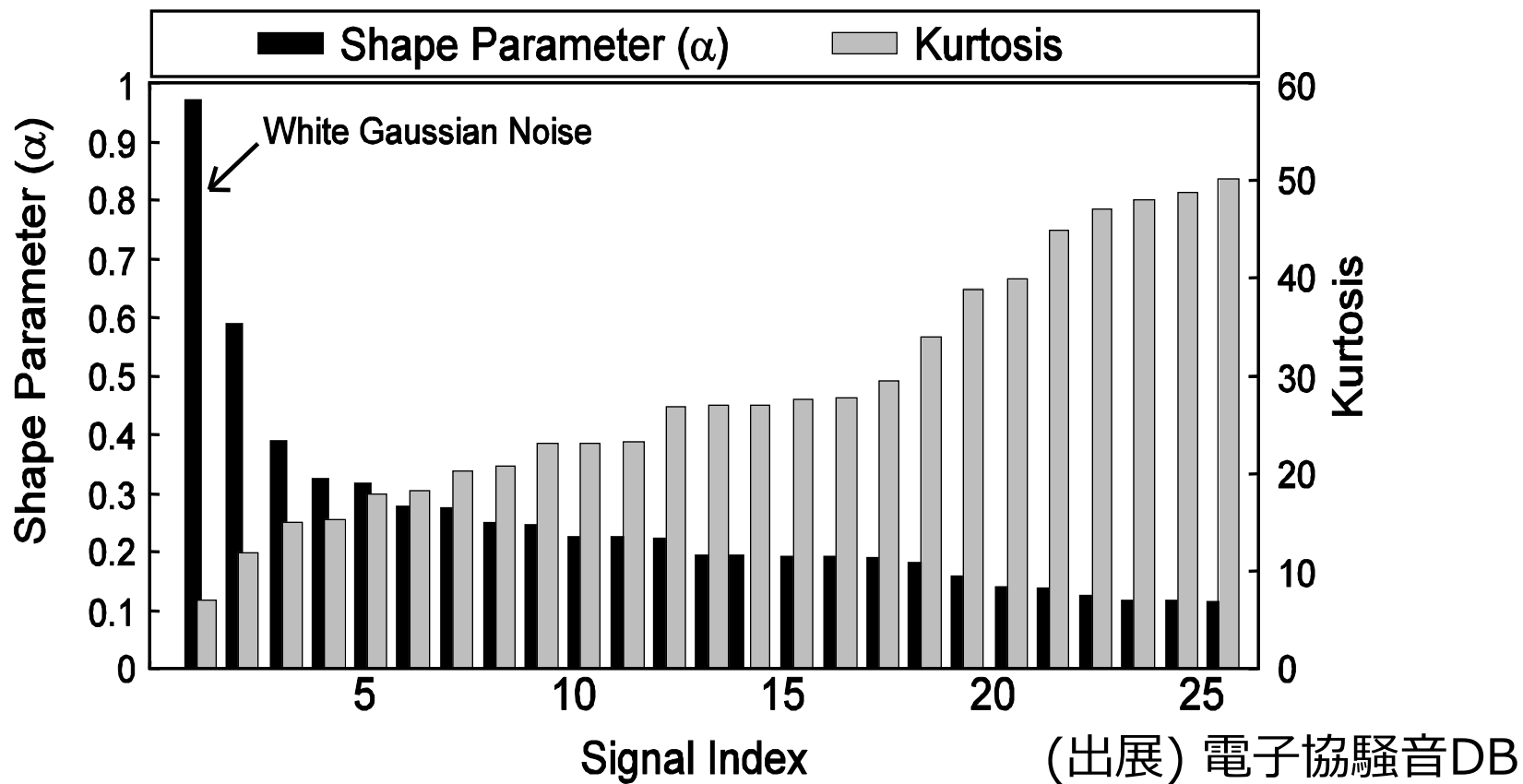
- 再帰的な性質など

## ■ モデリング例

- ・ ガウシアンノイズのパワースペクトルの場合  
ガンマ分布の形状母数が1の場合に相当する

※ 尺度母数は分布形状に関係しない

## ■ 実環境音の形状母数とカートシス



カートシス：およそ10~50, 形状母数：およそ0.1~0.6  
 形状母数が小さいほどカートシスは大きくなる

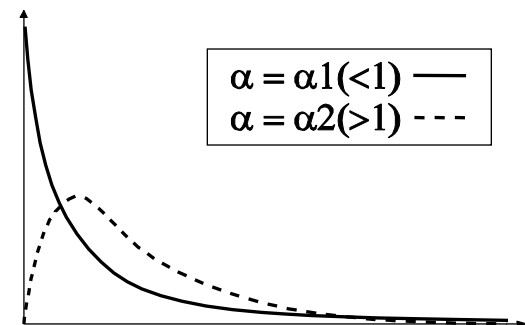
# ガンマ分布の母数推定

## ■ 母数推定

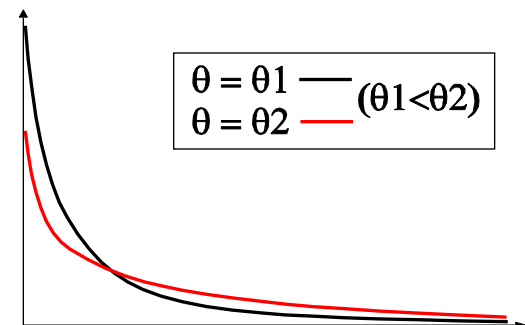
- ⇒ 生データから母数推定を行うことでモデル分布を得る
- ⇒ 形状母数と尺度母数を最尤推定法により推定

$$\left\{ \begin{array}{l} \hat{\alpha} = \frac{3 - \gamma + \sqrt{(\gamma - 3)^2 + 24\gamma}}{12\gamma}, \\ \hat{\theta} = \frac{E_l[x(l)]}{\hat{\alpha}}, \\ \gamma = \log ( E_l[x(l)] ) - E_l[ \log x(l) ], \end{array} \right.$$

$E_l[\cdot]$  は  $l$  に関する期待値演算子



形状母数と分布形状



尺度母数と分布形状

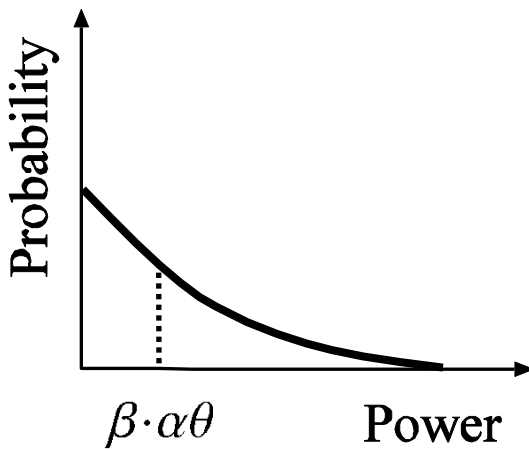
# SS前後の信号のモデリング

## ■ SSによる分布形状の変化を定式化

$$x \geq 0, \alpha > 0, \theta > 0,$$

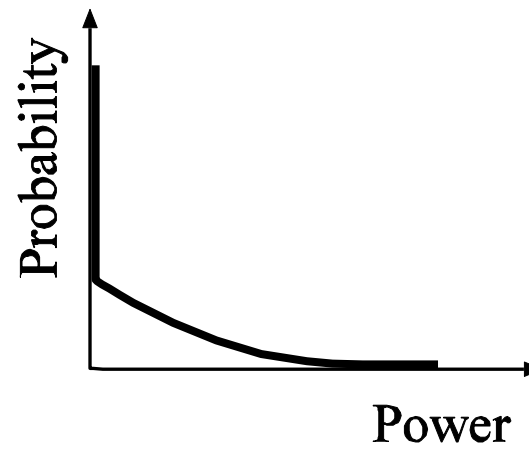
以下,  $C = 1/[\Gamma(\alpha)\theta^\alpha]$  とし,  $E[P(x)] = \alpha\theta$  を適用している

原信号の分布



$$P(x) = C \cdot x^{\alpha-1} e^{-\frac{x}{\theta}},$$

SS後の信号の分布



$$P(x) = \begin{cases} C (x + \beta\alpha\theta)^{\alpha-1} e^{-\frac{x+\beta\alpha\theta}{\theta}} & (x > 0), \\ C \int_0^{\beta\alpha\theta} x^{\alpha-1} e^{-\frac{x}{\theta}} dx & (x = 0). \end{cases}$$



# SS処理によるカートシス変化 (原信号のカートシス)

## ■ 原信号のカートシス

n次モーメントは,

$$\mu_n = \int_0^{\infty} x^n P(x) dx = \int_0^{\infty} C \cdot x^{(\alpha+n)-1} e^{-\frac{x}{\theta}} dx,$$

よって, カートシスの分子(4次モーメント)は

$$\mu_4 = \int_0^{\infty} C \cdot x^{(\alpha+4)-1} e^{-\frac{x}{\theta}} dx,$$

$X = \frac{x}{\theta}$  と変数変換すると,

$$\begin{aligned} \mu_4 &= C \int_0^{\infty} (X \cdot \theta)^{(\alpha+4)-1} e^{-X} \cdot \theta dX \\ &= C \theta^{\alpha+4} \int_0^{\infty} X^{(\alpha+4)-1} e^{-X} dX \\ &= C \theta^{\alpha+4} \cdot \Gamma(\alpha + 4). \end{aligned}$$

形状母数+モーメントの次数

$$\begin{aligned} x &\geq 0, \alpha > 0, \theta > 0, \\ C &= 1/[\Gamma(\alpha)\theta^\alpha], \\ P(x) &= C \cdot x^{\alpha-1} e^{-\frac{x}{\theta}}, \\ kurt &= \frac{\mu_4}{\mu_2^2}, \end{aligned}$$

# SS処理によるカートシス変化 (原信号のカートシス)

同様にして、カートシスの分母(2次モーメント)

$$\mu_2 = C\theta^{\alpha+2} \cdot \Gamma(\alpha + 2),$$

ゆえに、原信号のカートシス( $kurt_{orig}$ )は、

$$\begin{aligned} kurt_{orig} &= \frac{C\theta^{\alpha+4} \cdot \Gamma(\alpha + 4)}{[C\theta^{\alpha+2} \cdot \Gamma(\alpha + 2)]^2} \\ &= \frac{(\alpha + 3)(\alpha + 2) \cdot \Gamma(\alpha + 2)}{C\theta^{\alpha} \cdot \Gamma(\alpha + 2)^2} \\ &= \frac{(\alpha + 3)(\alpha + 2)}{\frac{1}{\Gamma(\alpha)\theta^{\alpha}} \cdot \theta^{\alpha}(\alpha + 1)\alpha \cdot \Gamma(\alpha)} \\ &= \frac{(\alpha + 3)(\alpha + 2)}{(\alpha + 1)\alpha}. \end{aligned}$$

$$\begin{aligned} x \geq 0, \alpha > 0, \theta > 0, \\ C &= 1/[\Gamma(\alpha)\theta^{\alpha}], \\ P(x) &= C \cdot x^{\alpha-1} e^{-\frac{x}{\theta}}, \\ kurt &= \frac{\mu_4}{\mu_2^2}, \end{aligned}$$

# SS処理によるカートシス変化 (SS後の信号のカートシス)

## ■ SS処理後のカートシス

n次モーメントは,

$$\mu_n = \int_0^{\infty} x^n [C(x + \beta\alpha\theta)^{\alpha-1} e^{-\frac{x + \beta\alpha\theta}{\theta}}] dx.$$

$$\begin{aligned} &x \geq 0, \alpha > 0, \theta > 0, \\ &C = 1/[\Gamma(\alpha)\theta^\alpha], \\ &P(x) = \begin{cases} C(x + \beta\alpha\theta)^{\alpha-1} e^{-\frac{x + \beta\alpha\theta}{\theta}} & (x > 0) \\ C \int_0^{\beta\alpha\theta} x^{\alpha-1} e^{-\frac{x}{\theta}} dx & (x = 0). \end{cases} \\ &kurt = \frac{\mu_4}{\mu_2^2}, \end{aligned}$$

テイラー展開により  $(x + \beta\alpha\theta)^{\alpha-1} = x^{\alpha-1} + \beta\alpha\theta(\alpha - 1)x^{\alpha-2} + \dots$ ,

カートシスの分子(4次モーメント)は,

$$\mu_4 = \int_0^{\infty} x^{(\alpha+4)-1} e^{-\frac{x + \beta\alpha\theta}{\theta}} dx + \beta\alpha\theta(\alpha - 1) \int_0^{\infty} x^{(\alpha+3)-1} e^{-\frac{x + \beta\alpha\theta}{\theta}} dx + \dots,$$

$X = \frac{x}{\theta}$  と変数変換し, 2次までで近似すると,

$$\mu_4 \approx C e^{-\alpha\beta} \left\{ \theta^{\alpha+4} \Gamma(\alpha+4) + \beta\alpha\theta(\alpha-1)\theta^{\alpha+3} \Gamma(\alpha+3) + \frac{(\beta\alpha\theta)^2}{2} (\alpha-1)(\alpha-2)\theta^{\alpha+2} \Gamma(\alpha+2) \right\}.$$

# SS処理によるカートシス変化 (SS後の信号のカートシス)

カートシスの分母(2次モーメント)は,  
0次で打ち切ったテイラー展開を利用して

$$\begin{aligned} \mu_2 &= \int_0^{\infty} x^2 \left[ C(x + \beta\alpha\theta)^{\alpha-1} e^{-\frac{x+\beta\alpha\theta}{\theta}} \right] dx \\ &\leq C e^{-\frac{\beta\alpha\theta}{\theta}} \Gamma(\alpha + 2). \end{aligned}$$

$$\begin{aligned} x &\geq 0, \alpha > 0, \theta > 0, \\ C &= 1/[\Gamma(\alpha)\theta^\alpha], \\ P(x) &= \begin{cases} C(x + \beta\alpha\theta)^{\alpha-1} e^{-\frac{x+\beta\alpha\theta}{\theta}} & (x > 0) \\ C \int_0^{\beta\alpha\theta} x^{\alpha-1} e^{-\frac{x}{\theta}} dx & (x = 0). \end{cases} \\ kurt &= \frac{\mu_4}{\mu_2^2}, \end{aligned}$$

カートシスは分布形状に依存する統計量で, 期待値  $\alpha\theta$  のスケールに依存しないため,  $\theta = 1/\alpha$  と正規化し簡単化する.  
以上より, SS後の信号のカートシス( $kurt_{ss}$ )に関して以下が成立する.

$$kurt_{ss} \geq \frac{e^{\alpha\beta}}{\alpha(\alpha+1)} \left\{ (\alpha+2)(\alpha+3) + \beta\alpha(\alpha+2)(\alpha-1) + \frac{(\beta\alpha)^2}{2}(\alpha-3)(\alpha-1) \right\}.$$

※分母を大きく見積もり, 全体で真の値より小さく見積もっている

また, 数値計算により

$$kurt_{ss} \geq kurt_{orig}$$

を確認した.

## ■ SS処理によるカートシス変化

SS後の信号のカートシスは指数部分が支配的であり、式の形から、変化量には比の対数が妥当と思われる。対数カートシス比(Log Kurtosis Ratio)を考えると、

$$\text{Log Kurtosis Ratio} = \log \left[ \frac{\text{kurt}_{ss}}{\text{kurt}_{orig}} \right] = \alpha\beta + \log \left\{ 1 + \frac{\beta\alpha(\alpha-1)}{(\alpha+3)} + \frac{(\beta\alpha)^2(\alpha-2)(\alpha-1)}{2(\alpha+2)(\alpha+3)} \right\}.$$

となる。これは

- 原信号の形状母数( $\alpha$ )
- 処理強度( $\beta$ )

のみの多項式からなる尺度である。

# 対数カートシス比

## ■ 対数カートシス比はミュージカルノイズ発生度合が

- 原信号の形状母数 ( $\alpha$ )
- 減算係数 ( $\beta$ )

に依存することを意味している

» (例) 形状母数 ( $\alpha$ ) を固定した場合

⇒ 対数カートシス比は減算係数にのみ依存し、減算係数が高いほどミュージカルノイズ発生度合も大きい

⇒ 経験則として一般的に知られている

» (例) 減算係数 ( $\beta$ ) を固定した場合

⇒ 対数カートシス比は原信号の形状母数にのみ依存し、原信号のPDF形状がなだらかなほど（形状母数 ( $\alpha$ ) が大きいほど）ミュージカルノイズ発生度合も大きい

⇒ 新たな発見

# 主観評価実験

- 目的：主観値(ミュージカルノイズスコア※)と客観値(対数カートシス比)の対応の調査

※ ミュージカルノイズスコア: ミュージカルノイズ発生度合の主観スコア (0:Natural, ..., 4:Harmful の5段階)

音源	信号長	10 [sec.]
	雑音	電子協騒音DB から4種
	音声	JNAS 4文(男女各2名)
SS条件	雑音推定	無音声区間の時間平均パワー
	減算係数	{0,0.4,0.8,1.2,1.6,2.0}
	フロアリング	0
評価	参照音	電子協騒音DBから5種とSSを適用したものの計10音源
	評価規範	ミュージカルノイズスコア※
	被験者	男性8名、女性1名

# 実験に使用した音源 減算量と対数カートシス比の関係

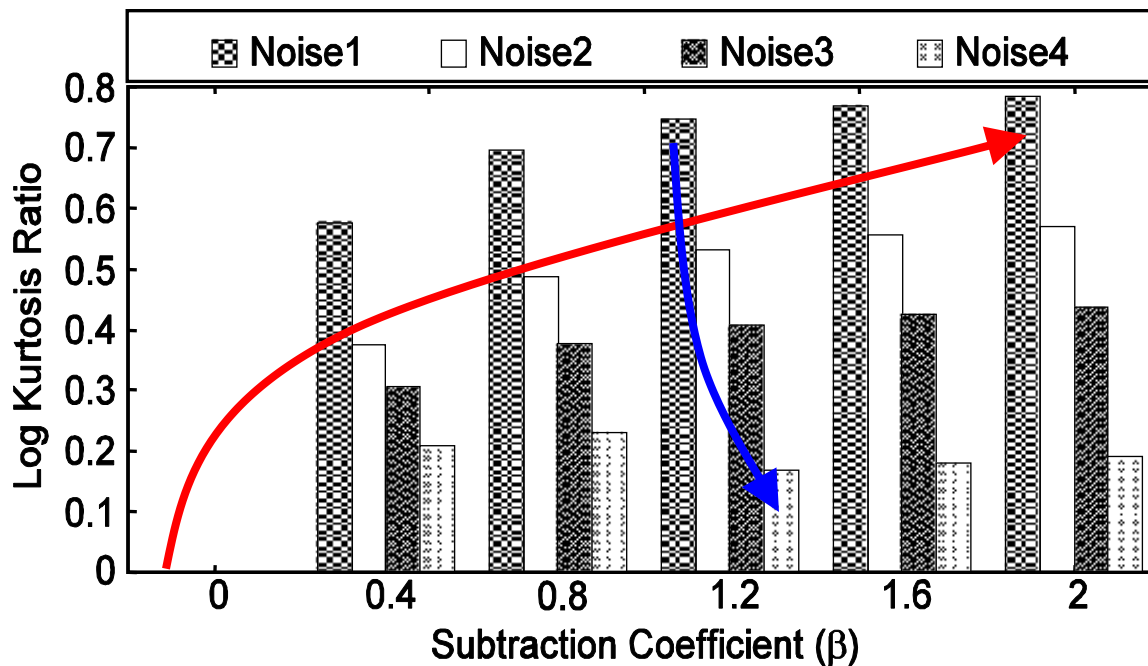
音源 / 環境 / カートシス / 形状母数

Noise1 / 駅 / 17 / 0.27

Noise2 / 人ごみ / 28 / 0.19

Noise3 / 展示会場 / 38 / 0.12

Noise4 / 病院 / 56 / 0.1



<音源固定>

減算係数：大→対数カートシス比：大

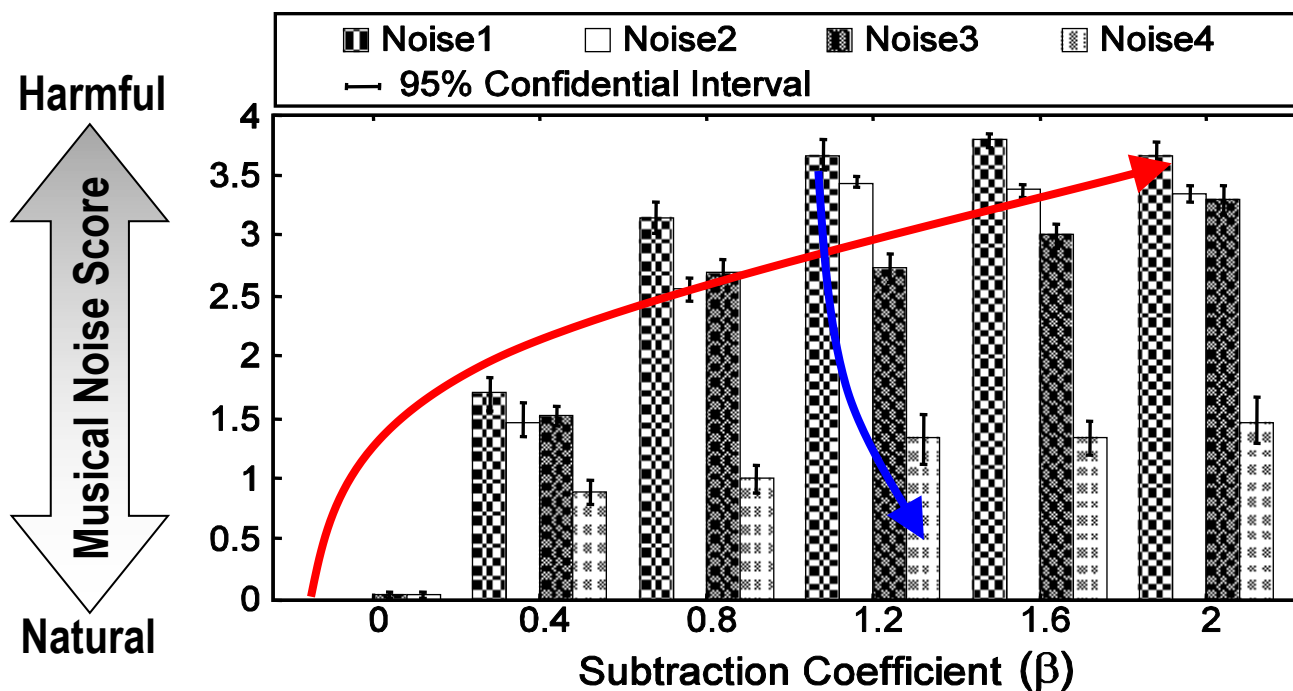
<減算係数固定>

原信号のカートシス：大→対数カートシス比：小



# 結果 (1/2)

## ■ 減算量, 音源とミュージカルノイズスコアの関係



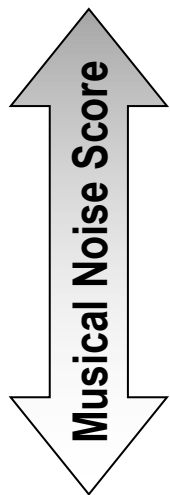
<音源固定> 減算係数 : 大 → ミュージカルノイズスコア : 大

<減算係数固定> 原信号のカートシス : 大 → ミュージカルノイズスコア : 小

# 結果 (2/2)

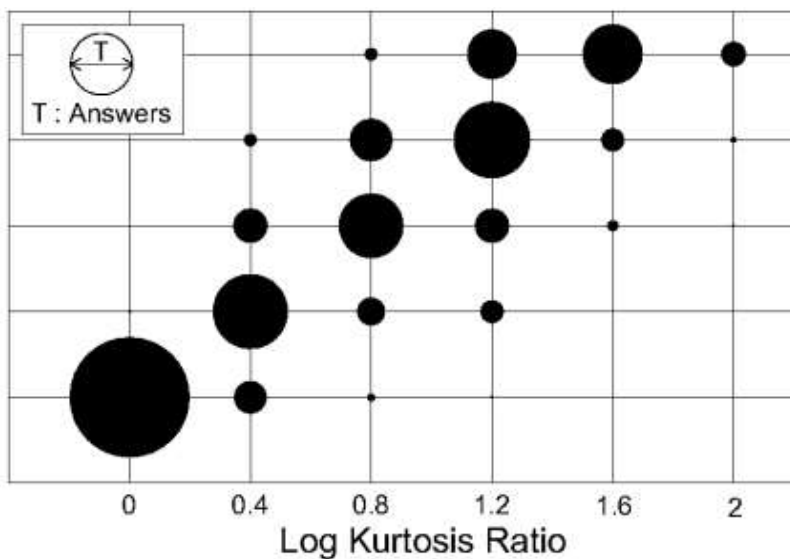
## ■ 対数カートシス比, 減算係数とミュージカルノイズスコア

Harmful

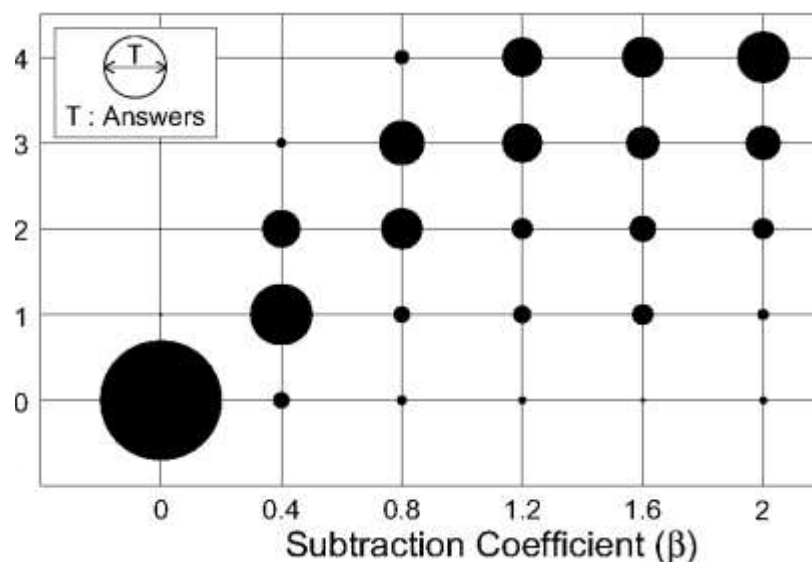


Natural

Correlation : 0.84



Correlation : 0.65



対数カートシス比は原信号の形状母数を勘案できる分相関が強い  
 ⇨ 原信号のPDF形状でミュージカルノイズ発生度合が異なる

# カートシス比 [Uemura, Saruwatari, 2008\*]

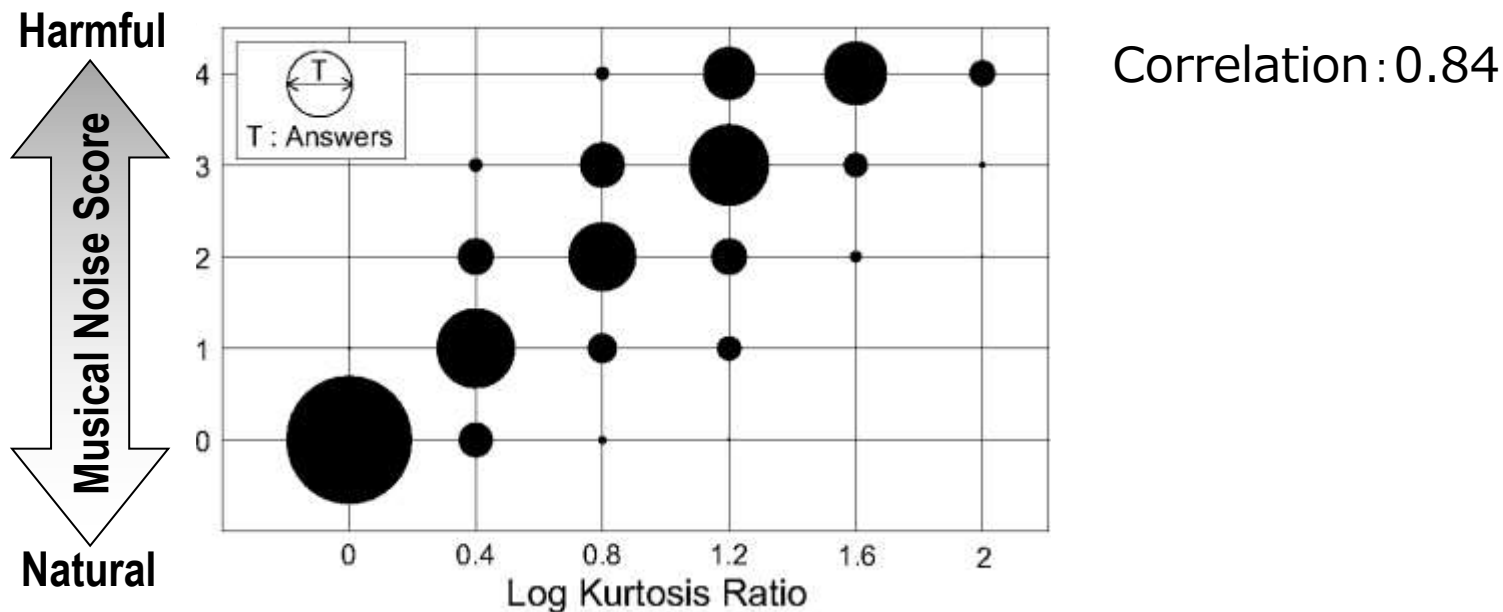
## ■ 処理前後のカートシス変化比（源信号尖度で正規化）

$$\text{kurtosis ratio} = \frac{\text{kurt}_{\text{proc}}}{\text{kurt}_{\text{org}}}$$

$\text{kurt}_{\text{org}}$  入力信号の尖度  
 $\text{kurt}_{\text{proc}}$  出力信号の尖度

カートシス比が1 ⇒ ミュージカルノイズ無し

カートシス比が1より大 ⇒ ミュージカルノイズ多い



[\*] Y. Uemura et al., "Automatic optimization scheme of spectral subtraction based on musical noise assessment via higher-order statistics," Proc. of IWAENC 2008,

# 雑音抑圧量 (Noise Reduction Ratio: NRR)

**NRR**

出力SN比－入力SN比

$$\text{NRR} = 10 \log_{10} \frac{E[s_{\text{out}}^2]/E[n_{\text{out}}^2]}{E[s_{\text{in}}^2]/E[n_{\text{in}}^2]}$$

処理後の音声信号:  $s_{\text{out}}$

処理後の音声信号:  $n_{\text{out}}$

処理前の音声信号:  $s_{\text{in}}$

処理前の雑音信号:  $n_{\text{in}}$

ここで信号処理における雑音成分の抑圧量は音声成分に影響を与えないと仮定すると

$$\text{NRR} = 10 \log_{10} \frac{E[n_{\text{in}}^2]}{E[n_{\text{out}}^2]} = 10 \log_{10} \frac{\mu'_1}{\mu_1}$$

処理前の1次モーメント:  $\mu_1$

処理後の1次モーメント:  $\mu'_1$

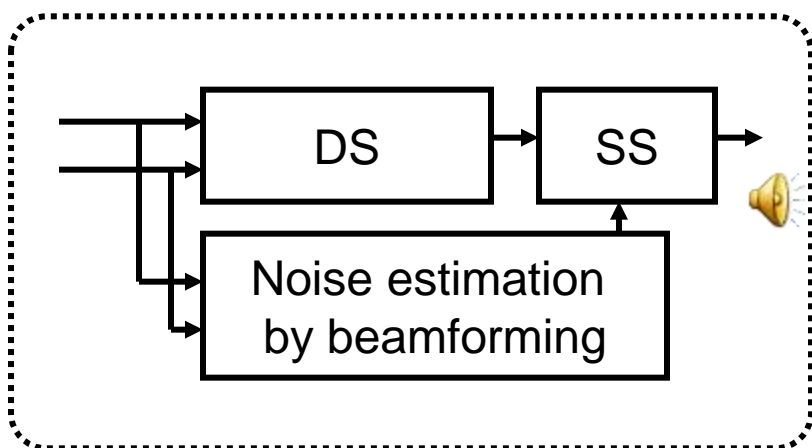
1次モーメントで**量**を、2次・4次モーメントで**質**を表す

# 高次統計量に基づく ミュージカルノイズ制御型アレー

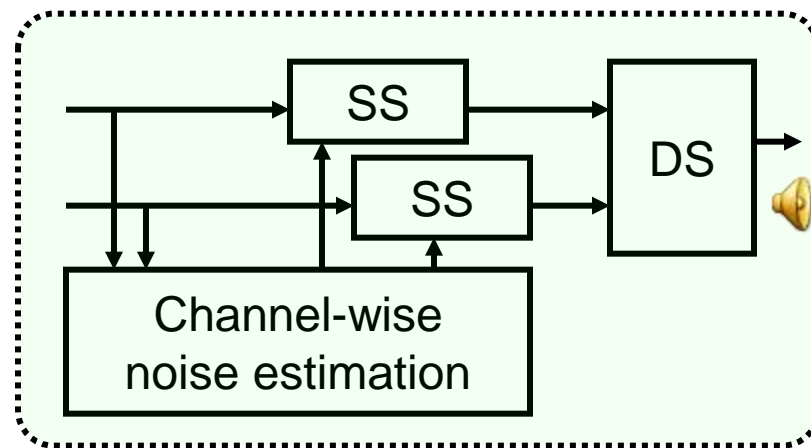
# ミュージカルノイズの発生量を抑える アレー信号処理と非線形処理の統合手法

## チャンネル毎スペクトル減算とアレー信号処理の組み合わせ[\*]

- チャンネル毎にスペクトル減算を行った後、遅延和アレー (DS) を適用する (下図右)
- 雑音抑圧性能を損なわず、ミュージカルノイズの発生を抑制できる。
- 尖度に基づく解析を行い、**ガウス性雑音に対しては**理論的にも尖度を低減、すなわちミュージカルノイズを低減できることを示す。



従来良く見られる組み合わせ方  
(BF+SS)

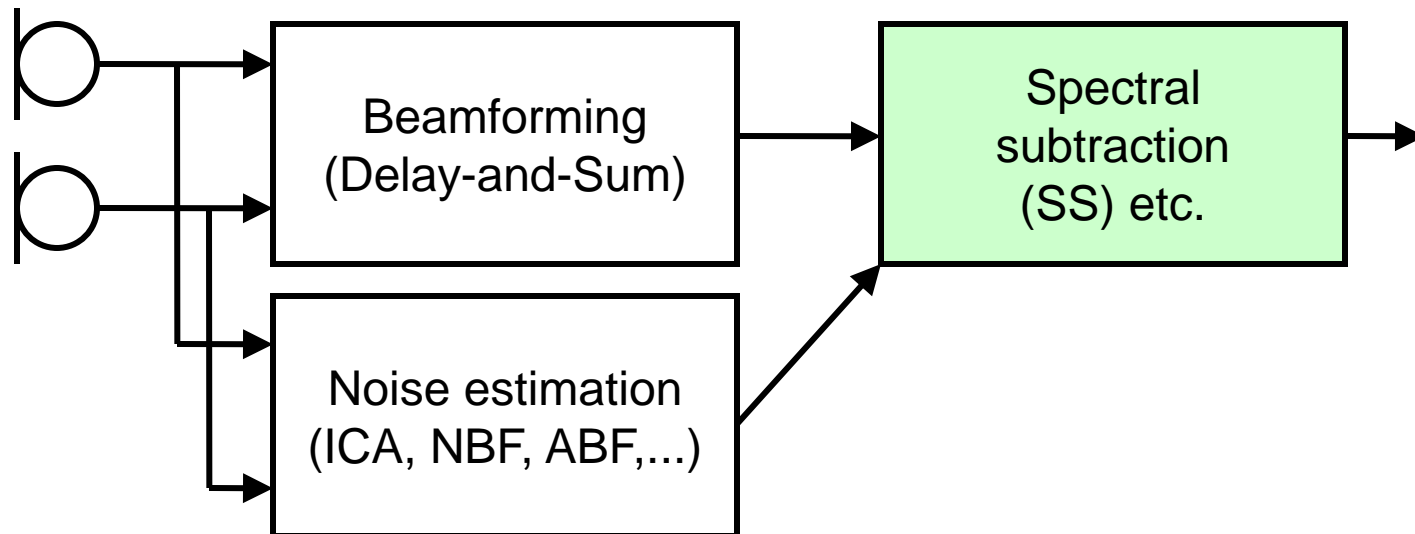


検討する「低ミュージカルノイズ」な  
組み合わせ方 (chSS+BF)

[\*] Takahashi, Saruwatari et al., "Musical-Noise Analysis in Methods of Integrating Microphone Array and Spectral Subtraction Based on Higher-Order Statistics," EURASIP JASP, vol. 2010, Article ID 431347, 2010.

# 従来のポストフィルタのつなぎ方(BF+SS)

ビームフォーミングの後ろに、ポスト音声強調系(スペクトル減算など)を接続

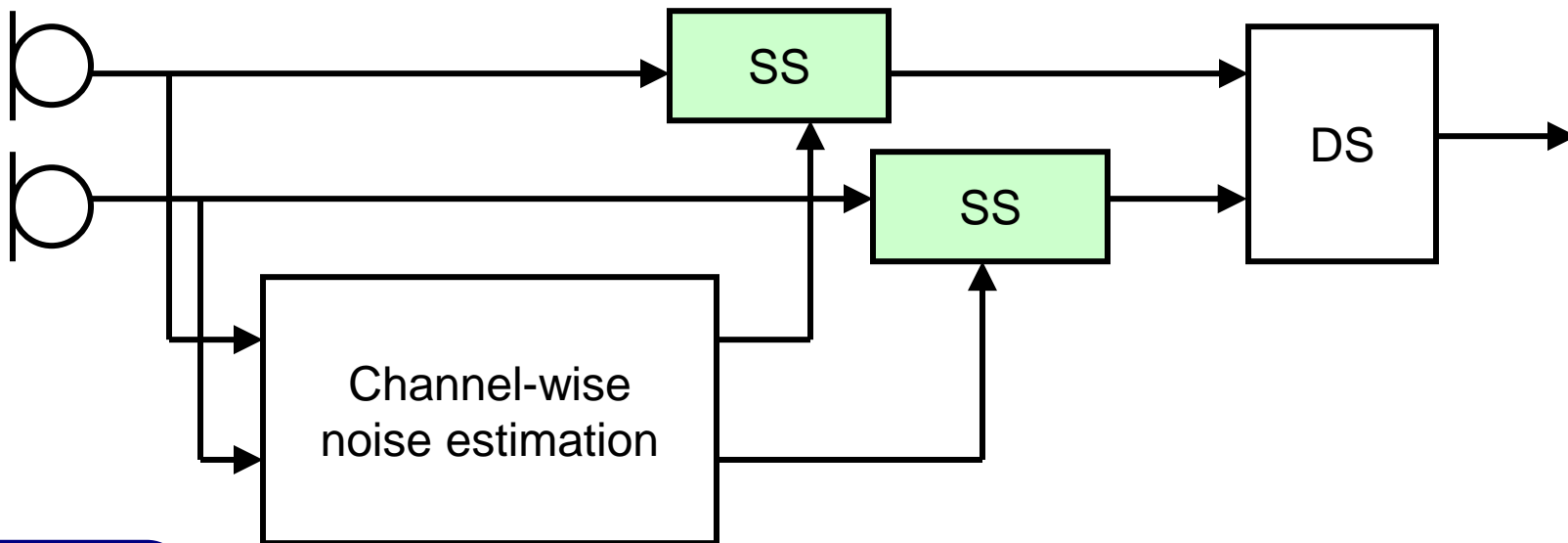


## 処理手順

1. ビームフォーミング (DS) により**シングルチャネル**の目的音強調信号を得る.
2. 同時に、死角制御型ビームフォーマ(NBF)や、適応アレー, ICA 等を用いて**シングルチャネル**の雑音推定信号を得る.
3. 最後に 1, 2 で得られた信号を基に統計的音声強調を行い雑音抑圧信号を得る.

# 提案法: ミュージカルノイズを低減可能な スペクトル減算のつなぎ方 (chSS+BF)

チャンネル毎にスペクトル減算を行ってから、後段でビームフォーミング



## 処理手順

1. チャンネル毎に雑音推定を行う。  
(この推定はたとえば, Single-input multiple-output ICA (SIMO-ICA)などで行う)
2. チャンネル毎にスペクトル減算を行い, マルチチャンネルの雑音抑圧信号を得る.
3. 2 で得られたマルチチャンネルの信号に対して, ビームフォーミングを行う.

この SS の接続方法だと処理後の尖度を下げることが可能 (詳細は後述)



# 尖度に基づく解析の指針

処理後の尖度の増加量が小さい=ミュージカルノイズの発生量が少ない

⇒ 尖度を基に解析を行い chSS+BF の尖度増加量が小さいことを示す

## 解析の流れ

1. まず, スペクトル減算および, DS によってどのくらい尖度が増加するかを示す.
2. 次に, 上記の解析結果を基に,  
chSS+BF および BF+SS によってどのくらい尖度が増加するかを示す.

## 仮定する信号

- $J$  チャンネルのマルチチャンネル信号  $x_j$  ( $j=1, \dots, J$ ) を考える.
- チャンネル毎に独立かつ同一分布に従う (i.i.d.) 信号が入力されているとする.
- また, **時間領域**における入力信号が従う分布の確率密度関数(p.d.f.) は両側対称かつ平均ゼロとする.
  - この仮定により, 1 次を除く奇数次モーメント/キュムラントはゼロになる.
- また, 入力信号は**パワースペクトル上**ではガンマ分布でモデリングできるとする.

- これらの仮定は入力信号がガウス性であることを仮定しない
- すなわち実環境におけるほとんど全ての雑音を対象にできる

# スペクトル減算後の尖度解析

## スペクトル減算後の尖度

スペクトル減算後の尖度は以下のように表現できる [Uemura, Saruwatari, 2008]

$$\text{kurt}_{SS} \geq \frac{e^{\alpha\beta}}{\alpha(\alpha+1)} \left\{ (\alpha+2)(\alpha+3) + \beta\alpha(\alpha+2)(\alpha-1) + \frac{(\beta\alpha)^2}{2}(\alpha-3)(\alpha-1) \right\}.$$

$\alpha$  入力信号をガンマ分布でモデリングしたときの形状母数  
 $\beta$  スペクトル減算における減算係数

- スペクトル減算の減算係数 ( $\beta$ ) を大きくすると尖度が上昇する
- 入力信号の分布の形状 ( $\alpha$ ) により, 処理後の尖度が異なる.



**スペクトル減算後の尖度は  
減算係数だけでなく, 処理前の信号の統計的性質にも依存する.**

# DS後の信号の尖度解析

モーメントとキュムラントを用いて解析を進め、DS後の尖度がどうなるか明らかにする。ここでは以下の性質と仮説を用いる。

## キュムラントに関する和の性質

$$\text{cum}_n(aX + bY) = a^n \text{cum}_n(X) + b^n \text{cum}_n(Y),$$

$\text{cum}_n(X)$  は確率変数  $X$  の  $n$  次キュムラント

## パワー領域信号

周波数領域におけるパワー領域信号は以下のように表現できる

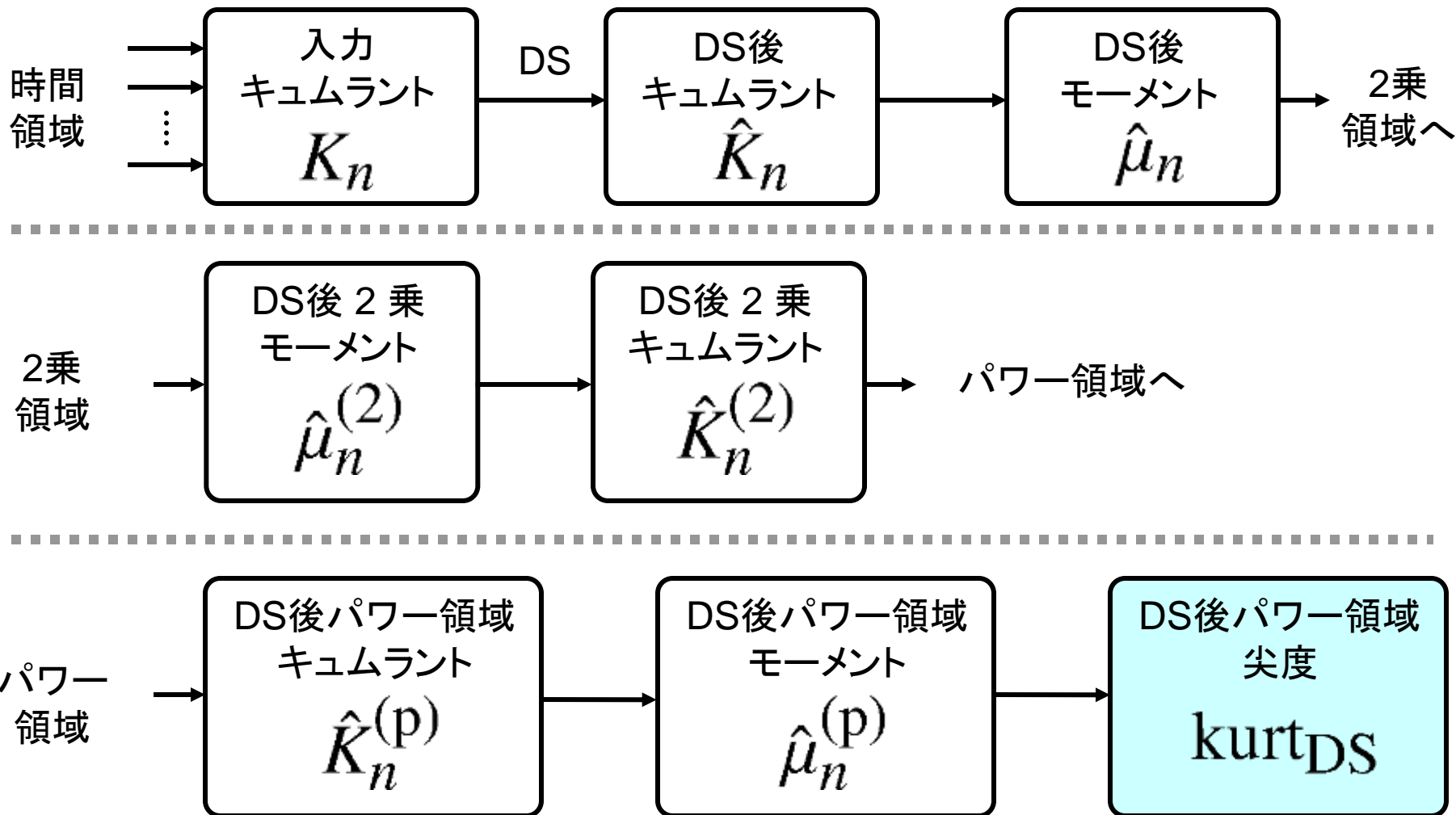
$$x_p = |x_{re} + i \cdot x_{im}|^2 = x_{re}^2 + x_{im}^2$$

$x_p$  パワー領域信号  
 $x_{re}$  実部  
 $x_{im}$  虚部

ここで、実部、虚部が互いに i.i.d. な信号であると仮定すると、パワー領域信号は2つの i.i.d. な信号の2乗信号の和になるため、キュムラントの和の性質からパワー領域のキュムラントを求めることができる。

# DS 後の尖度の解析手順

以下のように、時間領域から2乗領域を経て、パワー領域の尖度を求める



# DS 後の尖度

## DS後のパワー領域モーメント

$K_n$  入力信号の  $n$  次キュムラント

$$\hat{\mu}_2^{(p)} = 2 \left\{ \frac{1}{J^3} K_4 + \frac{4}{J^2} K_2^2 \right\}$$

$$\hat{\mu}_4^{(p)} = 2 \left\{ \frac{1}{J^7} K_8 + \frac{38}{J^6} K_4^2 + \frac{32}{J^6} K_6 K_2 + \frac{288}{J^5} K_4 K_2^2 + \frac{192}{J^4} K_2^4 \right\}$$

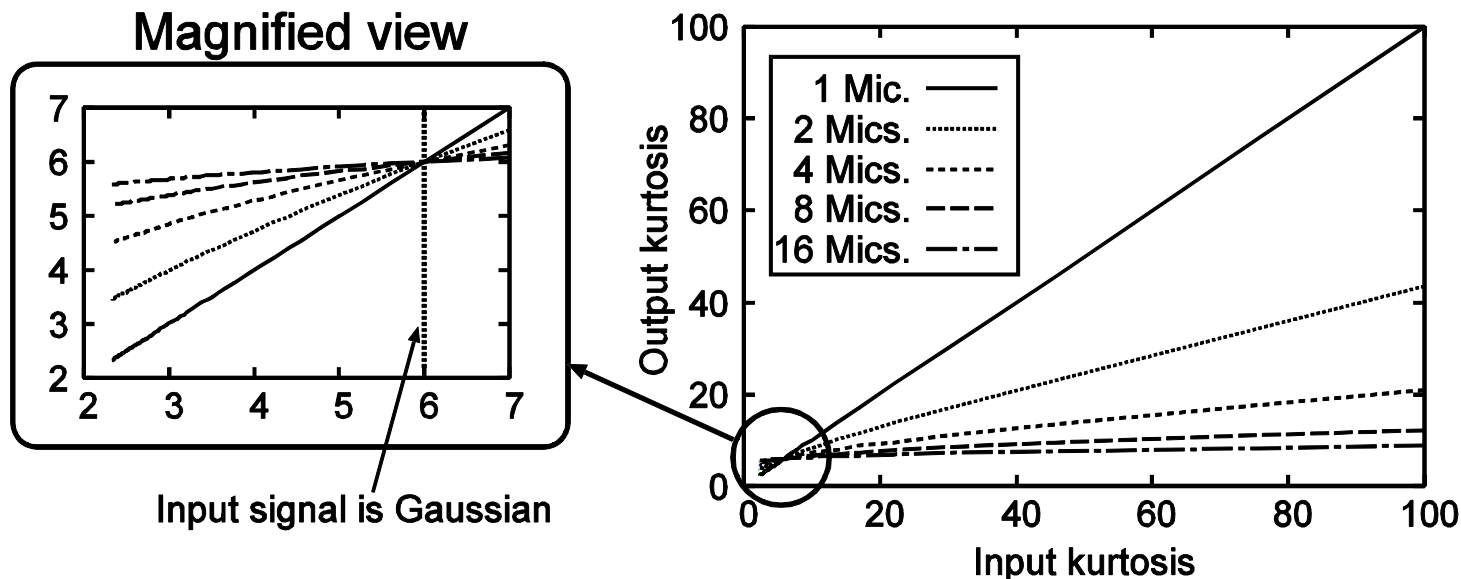
## DS後の尖度

$$\text{kurt}_{\text{DS}} = \frac{\hat{\mu}_2^{(p)}}{\hat{\mu}_4^{(p)}} = \frac{K_8 + 38JK_4 + 32JK_2K_6 + 288J^2K_2^2K_4 + 192J^3K_2^4}{2JK_4^2 + 16J^2K_2^2K_4 + 32J^3K_2^4}$$

ここで求めた DS 後の尖度は任意のキュムラントを設定でき  
あらゆる雑音に適用可能

# DS処理前後における尖度の関係

導出した尖度により, 以下のような DS の処理前後の尖度の関係が導かれる



- ここで, 尖度が6未満の信号は劣ガウシアン信号であり, 現実的な音響信号として存在することは稀であると考えられる.
- そこで入力信号の尖度が6以上であるとするとおおよそ以下のように近似できる.

$$\text{kurt}_{\text{DS}} \simeq J^{-1} \cdot (\text{kurt}_{\text{in}} - 6) + 6 \quad (\text{kurt}_{\text{in}} \geq 6)$$

# 最終出力の尖度比較

ここからは、BF+SS及びchSS+BF の最終出力の尖度を議論する。

## BF+SS の最終出力の尖度

$$\text{kurt}_{\text{BF+SS}} \geq \frac{e^{\hat{\alpha}\beta}}{\hat{\alpha}(\hat{\alpha}+1)} \left\{ (\hat{\alpha}+2)(\hat{\alpha}+3) + \beta\hat{\alpha}(\hat{\alpha}+2)(\hat{\alpha}-1) + \frac{(\beta\hat{\alpha})^2}{2}(\hat{\alpha}-3)(\hat{\alpha}-1) \right\}$$

$\hat{\alpha}$  はDS後の尖度に対応するガンマ分布の形状母数

## chSS+BFの最終出力の尖度

$$\text{kurt}_{\text{chSS+BF}} = J^{-1} \cdot (\text{kurt}_{\text{chSS}} - 6) + 6$$

$$\text{kurt}_{\text{chSS}} \geq \frac{e^{\tilde{\alpha}\beta}}{\tilde{\alpha}(\tilde{\alpha}+1)} \left\{ (\tilde{\alpha}+2)(\tilde{\alpha}+3) + \beta\tilde{\alpha}(\tilde{\alpha}+2)(\tilde{\alpha}-1) + \frac{(\beta\tilde{\alpha})^2}{2}(\tilde{\alpha}-3)(\tilde{\alpha}-1) \right\}$$

$\hat{\alpha}$  は入力信号の尖度に対応するガンマ分布の形状母数

# 最終出力の尖度比較

ここからは、BF+SS及びchSS+BF の最終出力の尖度を議論する。

## BF+SS の最終出力の尖度

$$\text{kurt}_{\text{BF+SS}} \geq \frac{e^{\hat{\alpha}\beta}}{\hat{\alpha}(\hat{\alpha}+1)} \left\{ (\hat{\alpha}+2)(\hat{\alpha}+3) + \beta\hat{\alpha}(\hat{\alpha}+2)(\hat{\alpha}-1) + \frac{(\beta\hat{\alpha})^2}{2}(\hat{\alpha}-3)(\hat{\alpha}-1) \right\}$$

$\hat{\alpha}$  はDS後の尖度に対応するガンマ分布の形状母数

## chSS+BFの最終出力の尖度

$$\text{kurt}_{\text{chSS+BF}} = J^{-1} \cdot (\text{kurt}_{\text{chSS}} - 6) + 6$$

$$\text{kurt}_{\text{chSS}} \geq \frac{e^{\tilde{\alpha}\beta}}{\tilde{\alpha}(\tilde{\alpha}+1)} \left\{ (\tilde{\alpha}+2)(\tilde{\alpha}+3) + \beta\tilde{\alpha}(\tilde{\alpha}+2)(\tilde{\alpha}-1) + \frac{(\beta\tilde{\alpha})^2}{2}(\tilde{\alpha}-3)(\tilde{\alpha}-1) \right\}$$

これらの最終出力の尖度のどちらが小さいかを議論する



# 最終出力の尖度の比較

## 比較指標

導出した BF+SS, chSS+BF の最終出力の尖度を比較するため、次の指標を用いる。

$$D = \text{kurt}_{\text{BF+SS}} - \text{kurt}_{\text{chSS+BF}}$$

- この  $D$  は出力尖度の二手法間での差を表している。
- $D > 0$  であれば、chSS+BF が BF+SS に比べて小さい尖度の信号を出力していることを示す。

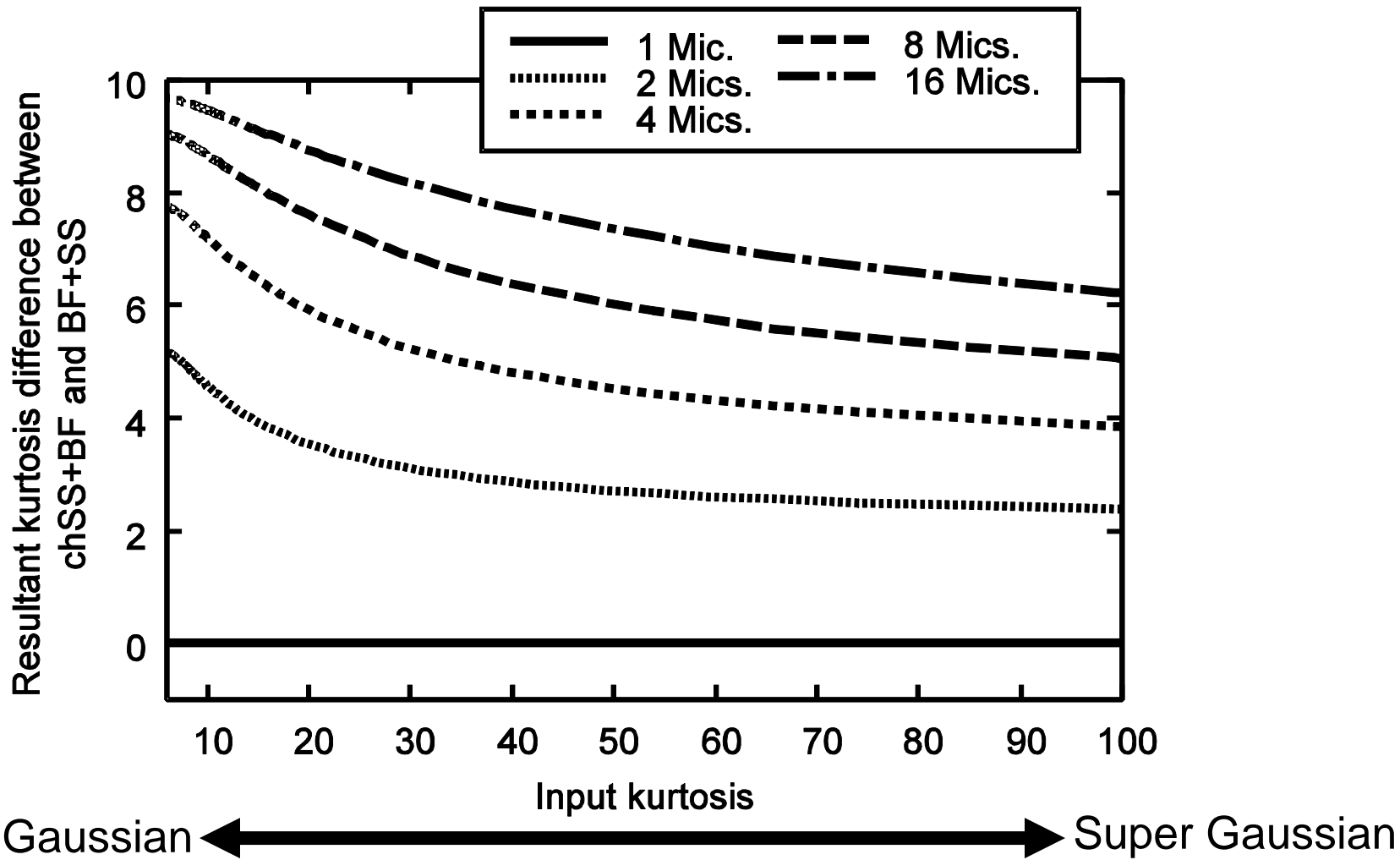


**$D > 0$  であれば、chSS+BF の方が  
ミュージカルノイズを低減できている事を示す。**

## 手順

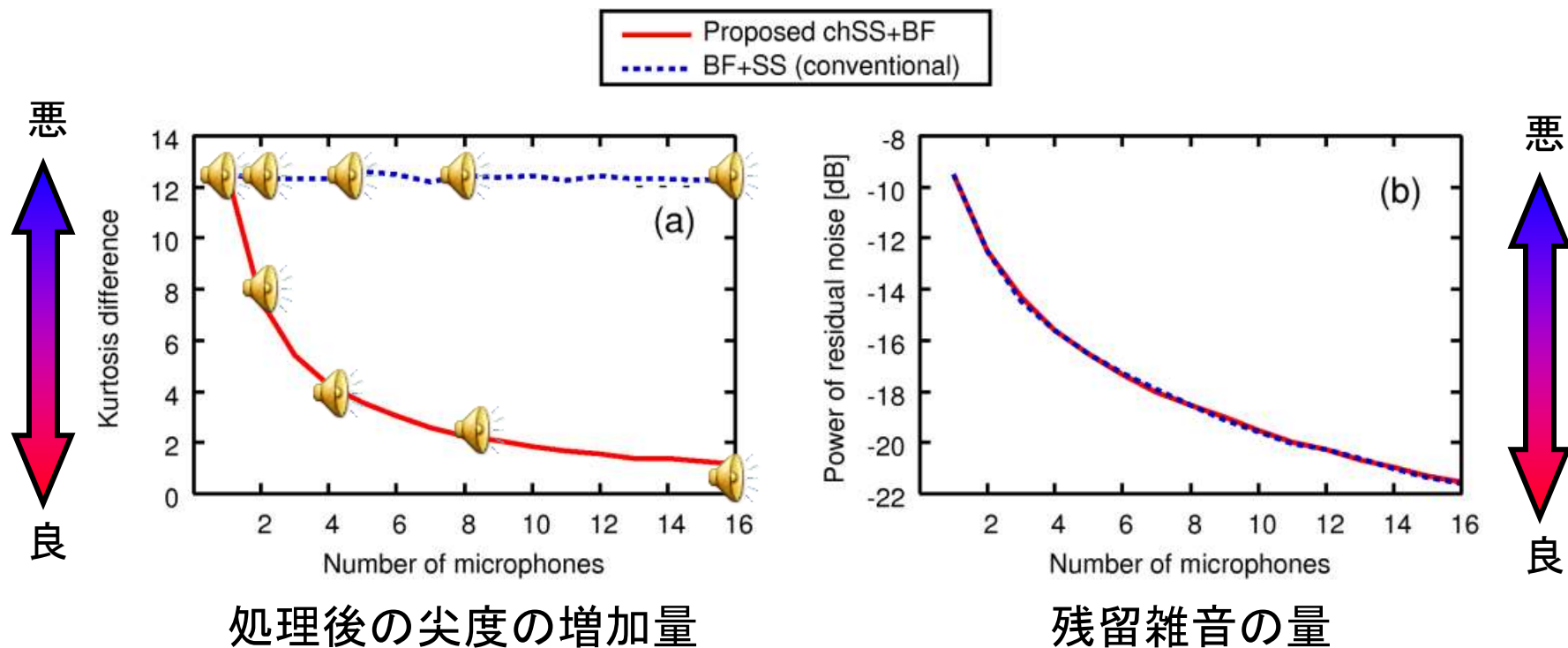
- 様々な尖度の入力信号に対する上記指標を観察する。
- なお、スペクトル減算の減算係数は 2.0 で固定。

# 最終出力の尖度の比較: 結果



■ chSS+BF は BF+SS に比べて、様々な入力信号に対して尖度を低減できる

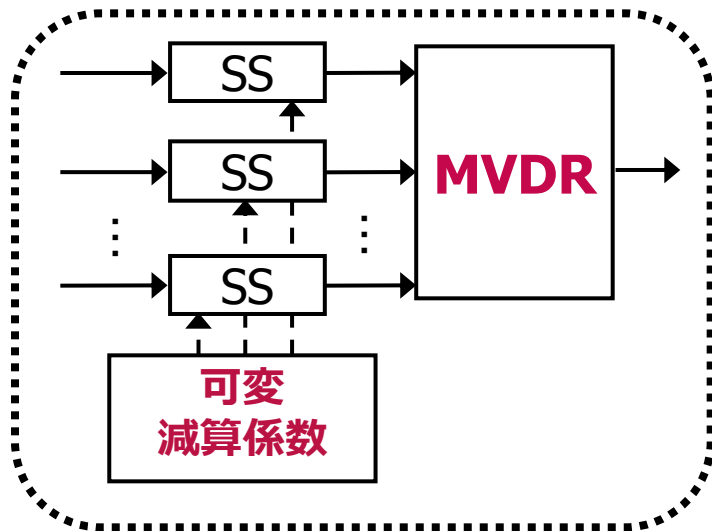
# 実験結果: 入力が高ウス性雑音の場合



- 雑音抑圧量は同じまま、ミュージカルノイズの発生を抑圧できる
- 尖度変化量とミュージカルノイズ発生量の関係の正当性を示唆

# ミュージカルノイズ制御型適応アレー

- ◆ **固定値の減算係数**：雑音環境によってミュージカルノイズ発生量が異なる  
→ 雑音環境に依らず一定のミュージカルノイズ発生量の出力が得られるように**減算係数を可変**にする
- ◆ **チャンネル毎SS+適応アレー (chSS+MVDR)** [Saruwatari, IEEE Trans. 2011]  
→ 拡散性雑音だけでなく点音源雑音にも適用できるようにchSS+DSのDS部を**MVDR**に拡張した「chSS+MVDR」を提案



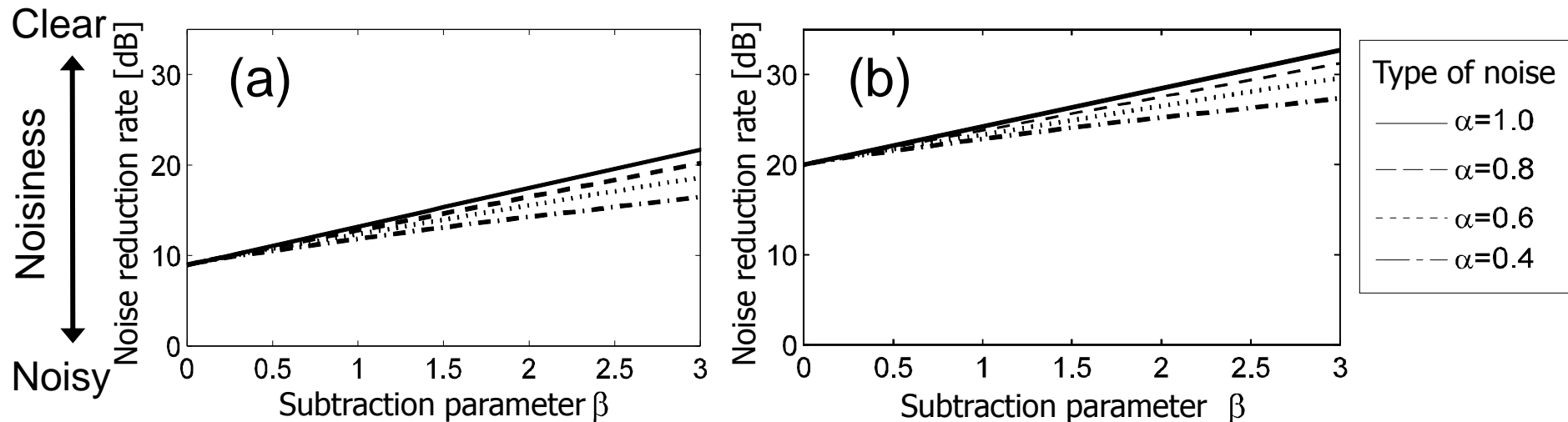
## MVDR (Minimum Variance Distortionless Response)

- ◆ 代表的な適応ビームフォーマの一つ
- ◆ **点音源雑音の抑圧性能が高い**

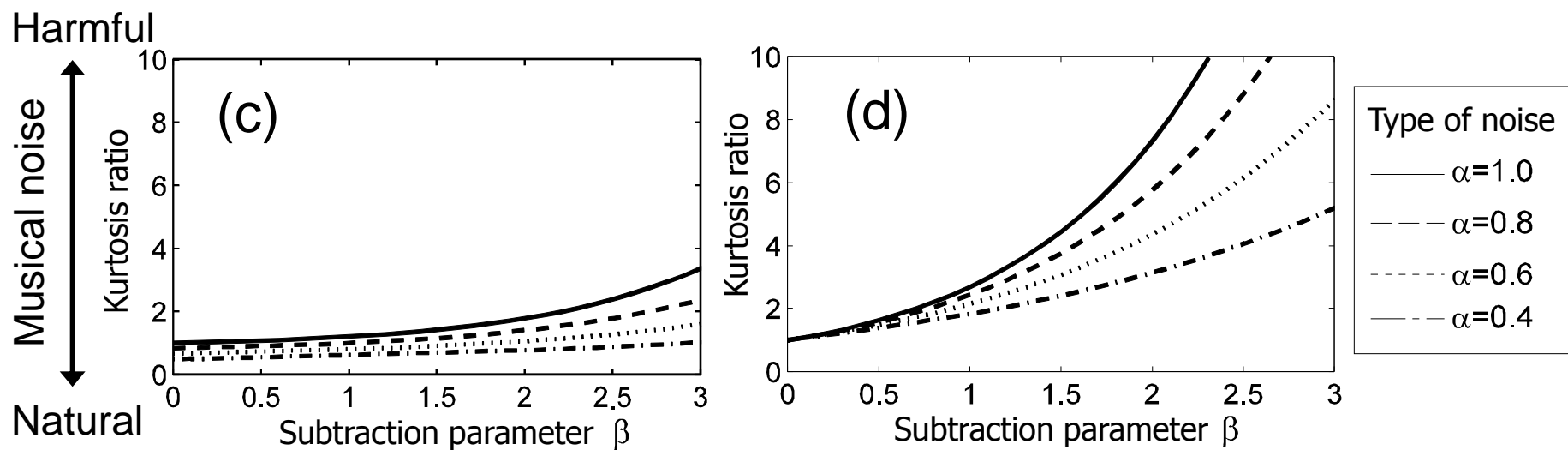
上記手法のミュージカルノイズ発生量とSNRを解析し、  
減算係数の決定方針を定める

# SNR改善量とカートシス比の解析結果

## ◆ SNR改善量

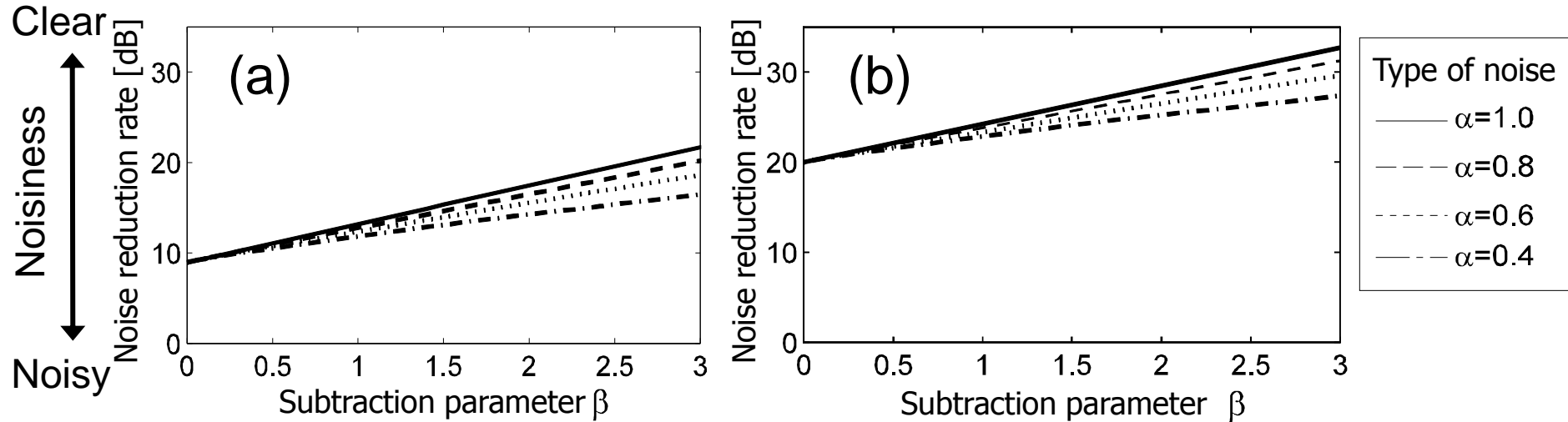


## ◆ カートシス比



# SNR改善量とカートシス比の解析結果

## ◆ SNR改善量



(a) 空間的無相関雑音の場合

MVDRのSNR改善性能が低く、SSの必要性は高い

(b) 点音源雑音の場合

MVDRのSNR改善性能が高く、SSの必要性は低い

Natural 0 0.5 1 1.5 2 2.5 3  
Subtraction parameter  $\beta$

0 0.5 1 1.5 2 2.5 3  
Subtraction parameter  $\beta$

# SNR改善量とカートシス比の解析結果

## ◆ SNR改善量

### (c) 空間的無相関雑音の場合

MVDRのミュージカルノイズ低減効果が大きく、  
ミュージカルノイズ発生量は少ない

### (d) 点音源雑音の場合

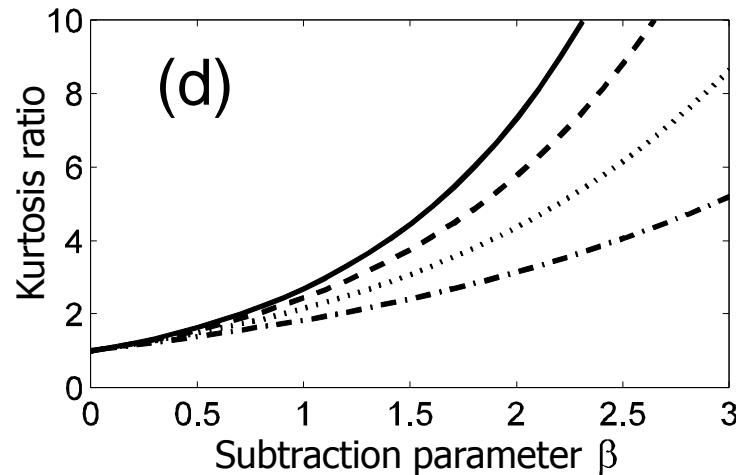
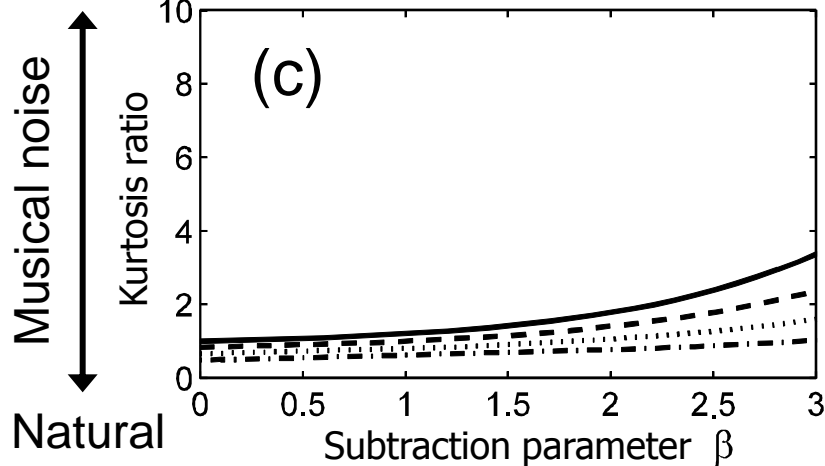
MVDRのミュージカルノイズ低減効果が小さく、  
ミュージカルノイズ発生量は多い

Subtraction parameter  $\beta$

Subtraction parameter  $\beta$

## ◆ カートシス比

Harmful



Type of noise

—  $\alpha=1.0$

- - -  $\alpha=0.8$

...  $\alpha=0.6$

- · -  $\alpha=0.4$

# chSS+MVDR の挙動解析のまとめ

## ◆ 空間的無相関な雑音の場合

(a) SNR改善性能：MVDR の性能が低く，SS に大きく頼ることになる

(c) ミュージカルノイズ発生量：比較的少ない

→ 許容される減算係数：大

## ◆ 点音源雑音の場合

(b) SNR改善性能：SS 処理をせずとも MVDR の性能が十分に高い

(d) ミュージカルノイズ発生量：比較的多い

→ 許容される減算係数：小



chSS+MVDR は，SS の減算係数を雑音毎に適切に設定さえすれば，拡散性/点音源の両雑音に対して SNR改善性能と音質性能をある程度両立できる



# 減算係数の自動決定アルゴリズムの提案

雑音の空間的特性や統計的性質に応じて、ユーザーが減算係数を事前に設定することは困難

→ **減算係数の自動決定アルゴリズムを提案**

[Step 0] 減算係数初期化

[Step 1] chSS の実行

減算係数 $\beta$ に基づいて  
チャンネル毎にSS を実行

[Step 2] MVDR の実行

chSS 出力信号を  
入力としてMVDR を実行

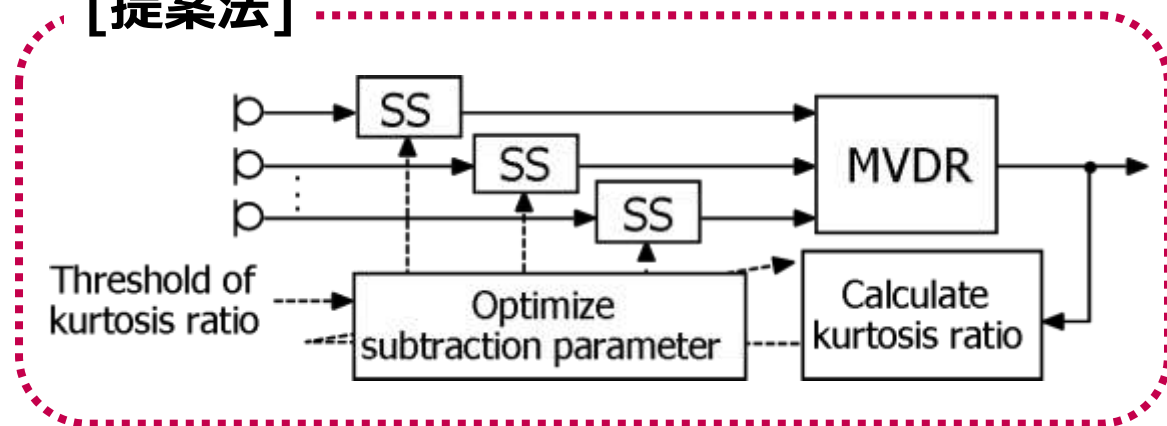
[Step 3] ミュージカルノイズ発生量の計量

システム全体によるカートシス比をサブバンド毎に算出

[Step 4] 減算係数 $\beta$  の更新

カートシス比の閾値を下回っていたら $\beta$ を $\Delta\beta$ だけ増加させ、  
上回っていたら $\beta$ を $\Delta\beta$ だけ減少させる。以上の操作をサブバンド毎に行う。  
更新された $\beta$ を用いて [Step 1] に進む。

[提案法]



# 減算係数の自動決定アルゴリズムの提案

雑音の空間的特性や統計的性質に応じて、ユーザーが減算係数を事前に設定することは困難

→ **減算係数の自動決定アルゴリズムを提案**

[Step 0] 減算係数初期化

[Step 1] chSS の実行

減算係数 $\beta$ に基づいて  
チャンネル毎にSS を実行

[Step 2] MVDR の実行

chSS 出力信号を  
入力としてMVDR を実行

[Step 3] ミュージカルノイズ発生量の計量

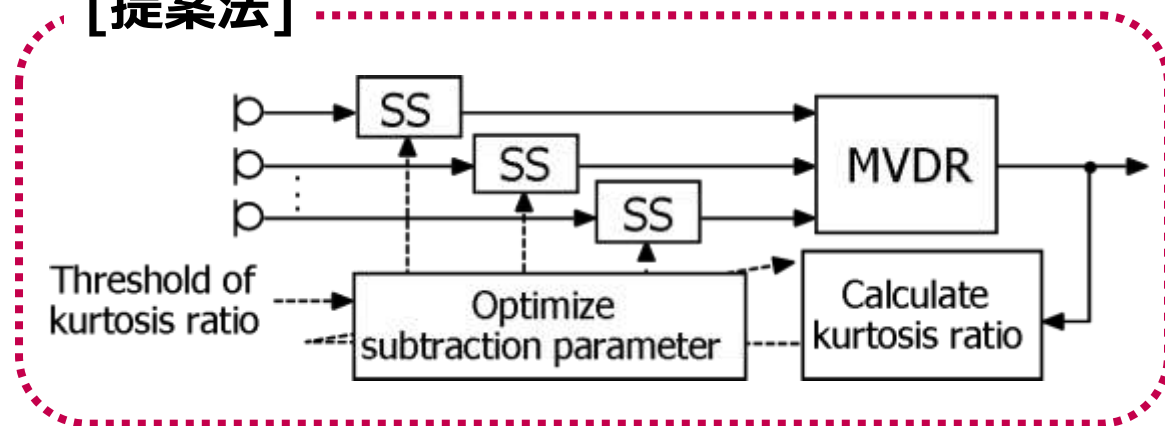
システム全体によるカートシス比をサブバンド毎に算出

[Step 4] 減算係数 $\beta$  の更新

許容のカートシス比に留める条件のもとで、減算係数を最大化

→ 許容の音質でSNRを自動的に最大化することが可能

[提案法]

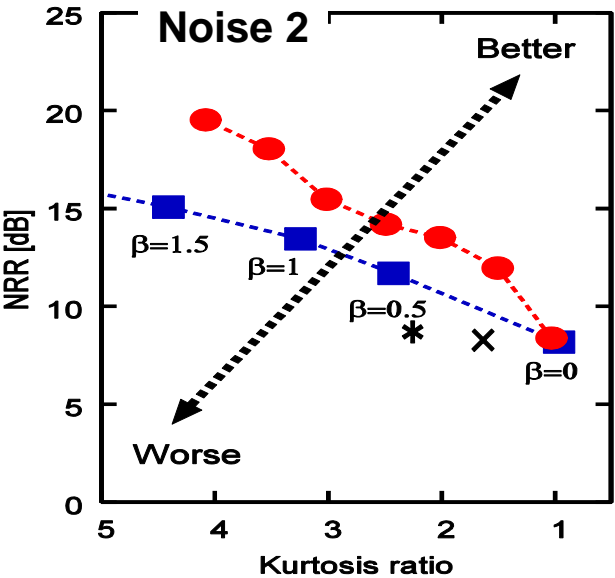
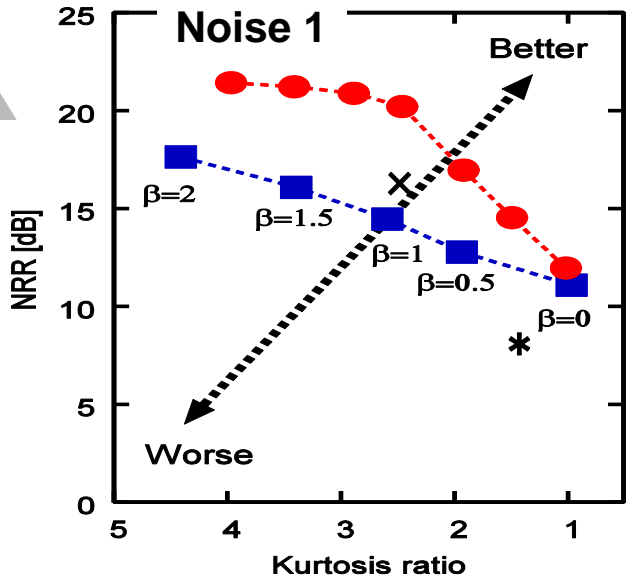


MVDR+SS ■ Proposed ● Zelinski PF \* , McCowan PF ×

$\beta$ : Subtraction strength of SS

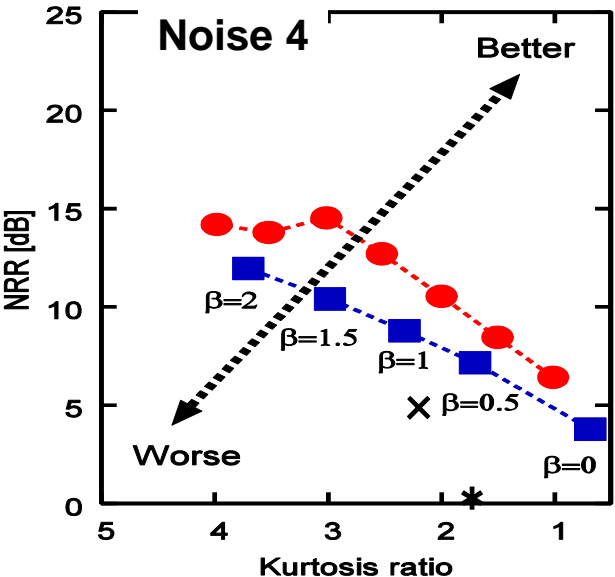
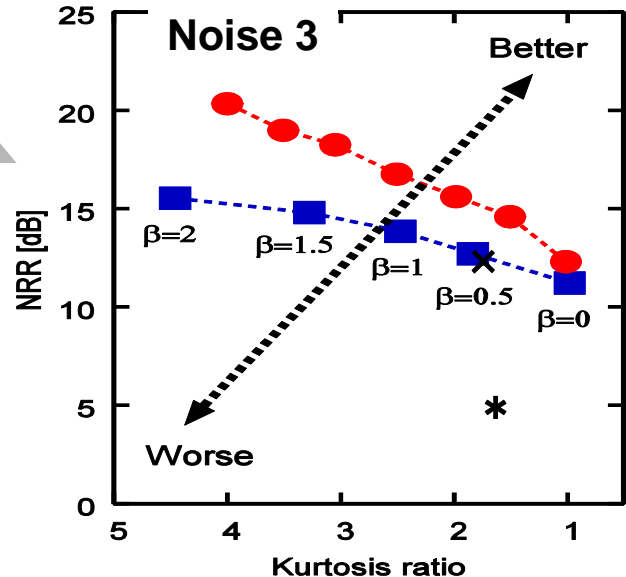
SNR

SNR



Noise 1: small room

Noise 2: office



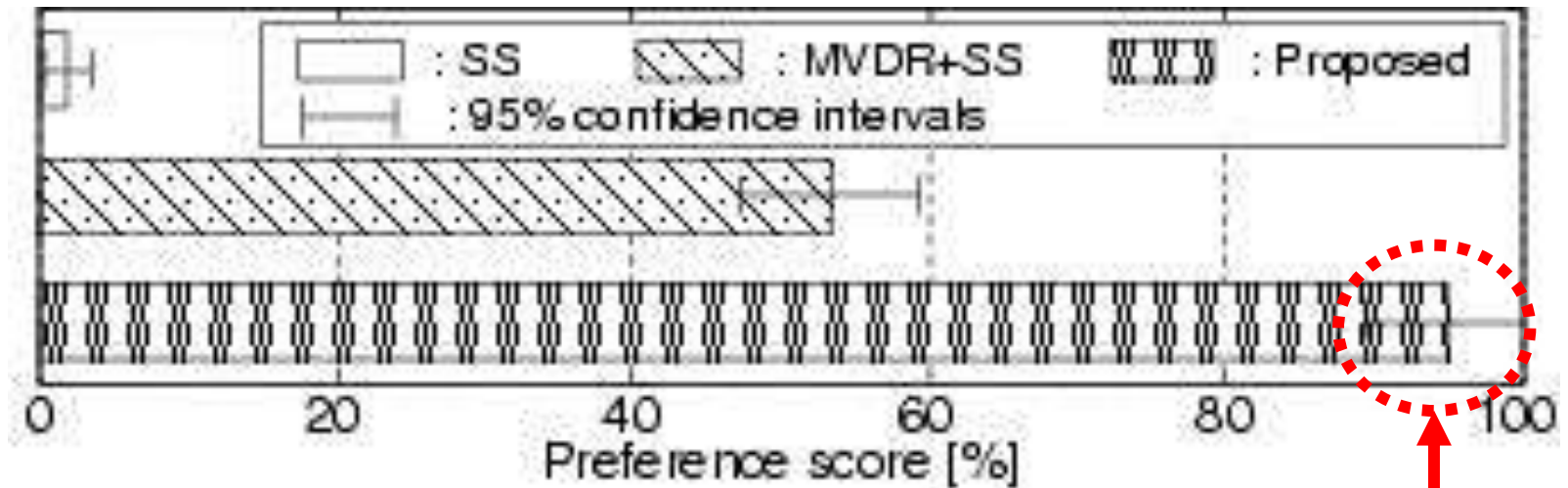
Noise 3: railway station

Noise 4: entrance hall of bldg

Less musical noise



# 主観評価実験結果



ほとんど全ての評価受聴者が、提案するミュージカルノイズ制御型アレーの音質を好むことが分かった。

# 高次統計量空間における不動点の発見と ミュージカルノイズフリー信号処理

# 各種信号処理における高次モーメント関数

## 一般化スペクトル減算 [EURASIP JASP 2010]

$$\mu_m^{\text{SS}} = \theta^m \mathcal{M}(\alpha, \beta, \eta, m)$$

$$\mathcal{M}(\alpha, \beta, \eta, m) = \frac{1}{\Gamma(\alpha)} \sum_{l=0}^m (-\beta\alpha)^l \frac{\Gamma(m+1)}{\Gamma(l+1)\Gamma(m-l+1)} \Gamma(\alpha+m-l, \beta\alpha) + \eta^{2m} \frac{\gamma(\alpha+m, \beta\alpha)}{\Gamma(\alpha)}$$

$\gamma(\alpha, z) = \int_0^z t^{\alpha-1} \exp(-t) dt$  : 第1種不完全ガンマ関数

$\Gamma(\alpha, z) = \int_z^\infty t^{\alpha-1} \exp(-t) dt$  : 第2種不完全ガンマ関数

## 一般化ウィーナフィルタ [IEEE Trans. ASLP 2011]

$$\mathcal{M}_{\text{QPWF}}(\alpha, \beta, m, \xi, \eta) = \frac{1}{\Gamma(\alpha)} \int_0^\infty \frac{t^{(\xi\eta+1)m+\alpha-1}}{\left\{ t^{\frac{\xi}{2}} + \beta \frac{\Gamma(\alpha+\frac{\xi}{2})}{\Gamma(\alpha)} \right\}^{2m\eta}} \exp(-t) dt$$

# 各種信号処理における高次モーメント関数

## ベイジアンMMSE音声振幅推定 [IWAENC2012]

$$\begin{cases} \mu_1 = K_1(|Y(f, \tau)|^2) \\ \mu_2 = K_2(|Y(f, \tau)|^2) + K_1^2(|Y(f, \tau)|^2) \\ \mu_4 = K_4(|Y(f, \tau)|^2) + 4K_3(|Y(f, \tau)|^2)K_1(|Y(f, \tau)|^2) \\ \quad + 3K_2^2(|Y(f, \tau)|^2) + 6K_2(|Y(f, \tau)|^2)K_1^2(|Y(f, \tau)|^2) \\ \quad + K_1^4(|Y(f, \tau)|^2) \end{cases}$$

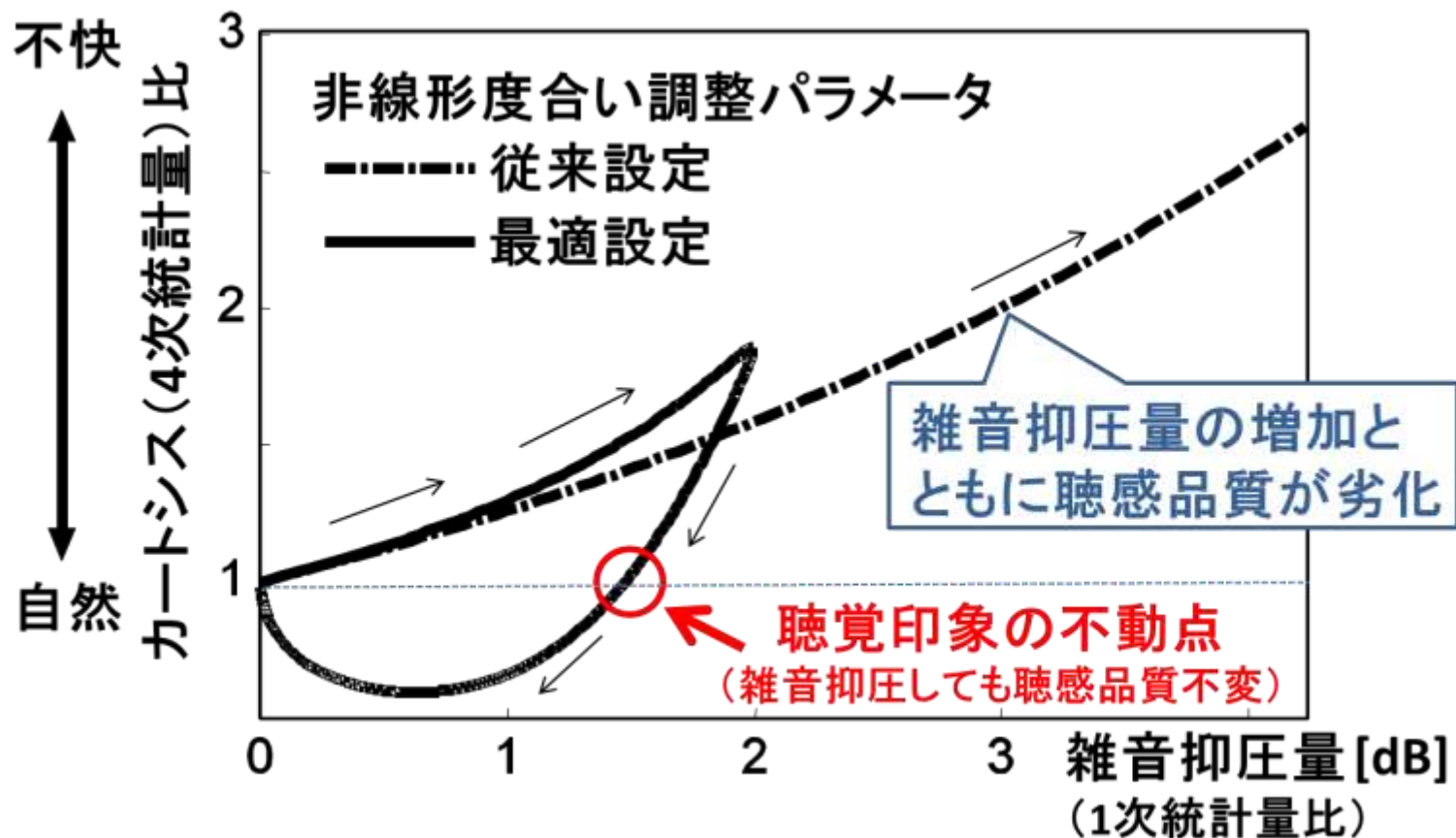
where  $K_m(|Y(f, \tau)|^2) =$

$$\sum_{i=1}^2 (-1)^{i-1} (1-\alpha)^{im} (R(\beta)\eta\theta)^m Z_i(m) K_m((\xi^{ml}(f, \tau))^i),$$

$$\begin{cases} K_1((\xi^{ml}(f, \tau))^i) = \mu_{i[SS]} \\ K_2((\xi^{ml}(f, \tau))^i) = \mu_{2i[SS]} - \mu_{i[SS]}^2 \\ K_3((\xi^{ml}(f, \tau))^i) = \mu_{3i[SS]} - 3\mu_{2i[SS]}\mu_{i[SS]} + 2\mu_{i[SS]}^3 \\ K_4((\xi^{ml}(f, \tau))^i) = \mu_{4i[SS]} - 4\mu_{3i[SS]}\mu_{i[SS]} - 3\mu_{2i[SS]}^2 \\ \quad + 12\mu_{2i[SS]}\mu_{i[SS]}^2 - 6\mu_{i[SS]}^4 \end{cases}$$

$$\mu_{m[SS]} = \sum_{l=0}^m (-\eta)^l \frac{\Gamma(m+1)\Gamma(\eta+m-l, \eta)}{\Gamma(\eta)\Gamma(l+1)\Gamma(m-l+1)}$$

# 高次統計量空間での不動点の発見

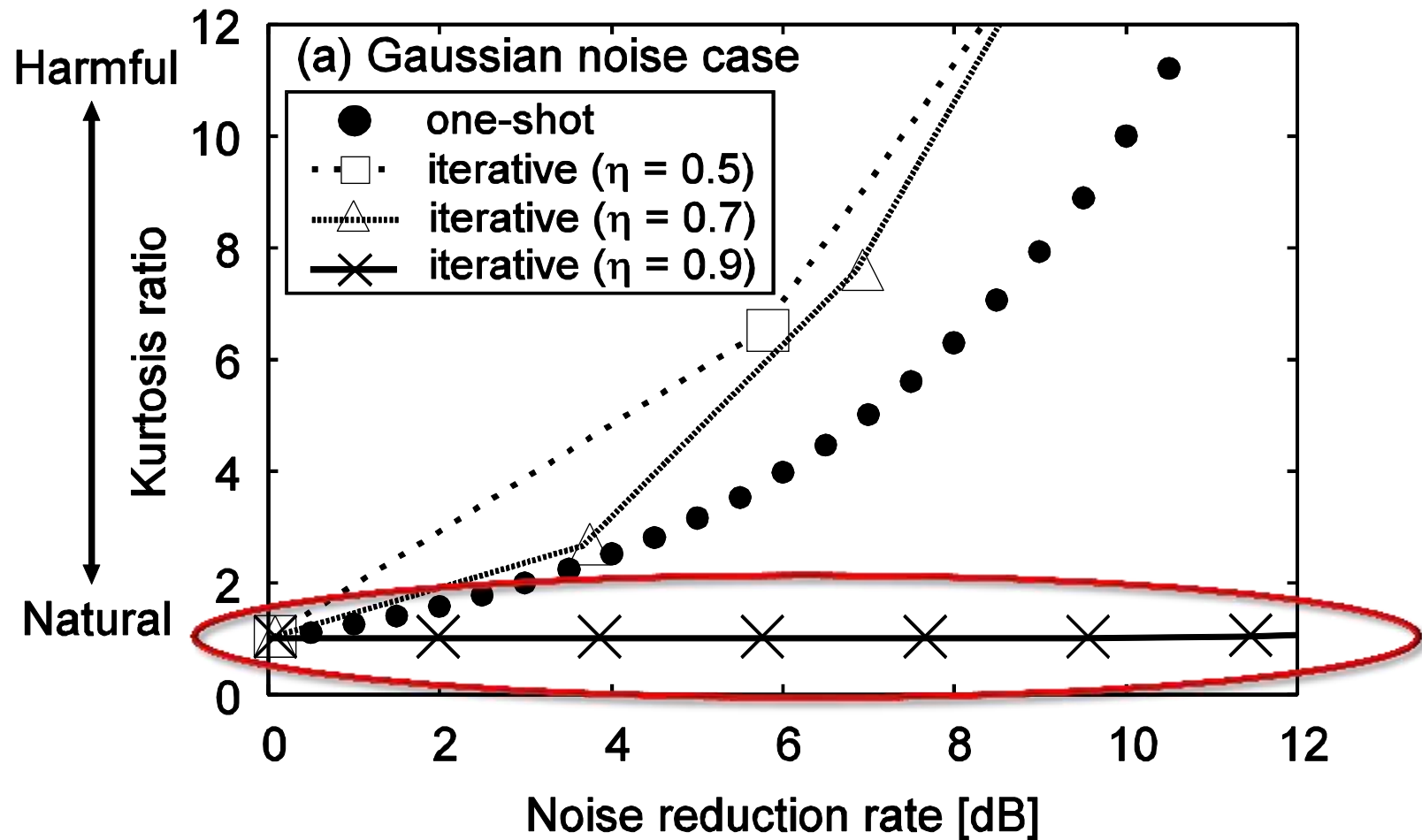


## ミュージカルノイズフリー信号処理 [Miyazaki, Saruwatari, IEEE Trans. ASLP 2012]

- カートシス比不動点の存在は聴覚印象の不動点を表す
- 一次統計量の増分が少ない場合は本処理を繰り返せば良い



# スペクトル減算を繰り返すと？



$\alpha = 1.0, \beta = 2.4, \eta = 0.9$  でミュージカルノイズフリー状態を達成→以降, 数学的に解を求める

# カートシス比不動点の解析例 (スペクトル減算法)

$$\text{kurt}(\alpha_0, \beta, \eta) = \frac{\mathcal{S}(\alpha_0, \beta, 4) + \eta^8 \mathcal{F}(\alpha_0, \beta, 4)}{(\mathcal{S}(\alpha_0, \beta, 2) + \eta^4 \mathcal{F}(\alpha_0, \beta, 2))^2} = \frac{(\alpha_0 + 3)(\alpha_0 + 2)}{(\alpha_0 + 1)\alpha_0}$$

処理後のカートシス

観測信号のカートシス

$$\mathcal{S}(\alpha, \beta, m) = \frac{1}{\Gamma(\alpha)} \sum_{l=0}^m (-\beta\alpha)^l \frac{\Gamma(m+1)}{\Gamma(l+1)\Gamma(m-l+1)} \cdot \Gamma(\alpha+m-l, \beta\alpha)$$

$$\mathcal{F}(\alpha, \beta, m) = \frac{\gamma(\alpha+m, \beta\alpha)}{\Gamma(\alpha)}$$

$\mathcal{H} = \eta^4$  と置くことで、2次方程式に帰着

$$\mathcal{H} = \{\mathcal{F}(\alpha_0, \beta, 4)(\alpha_0+1)\alpha_0 - \mathcal{F}^2(\alpha_0, \beta, 2)(\alpha_0+3)(\alpha_0+2)\}^{-1}$$

$$\left[ \mathcal{S}(\alpha_0, \beta, 2)\mathcal{F}(\alpha_0, \beta, 2)(\alpha_0+3)(\alpha_0+2) \right.$$

$$\pm \left[ \{\mathcal{S}(\alpha_0, \beta, 2)\mathcal{F}(\alpha_0, \beta, 2)(\alpha_0+3)(\alpha_0+2)\}^2 \right.$$

$$\left. - \{\mathcal{F}(\alpha_0, \beta, 4)(\alpha_0+1)\alpha_0 - \mathcal{F}^2(\alpha_0, \beta, 2)(\alpha_0+3)(\alpha_0+2)\} \right]$$

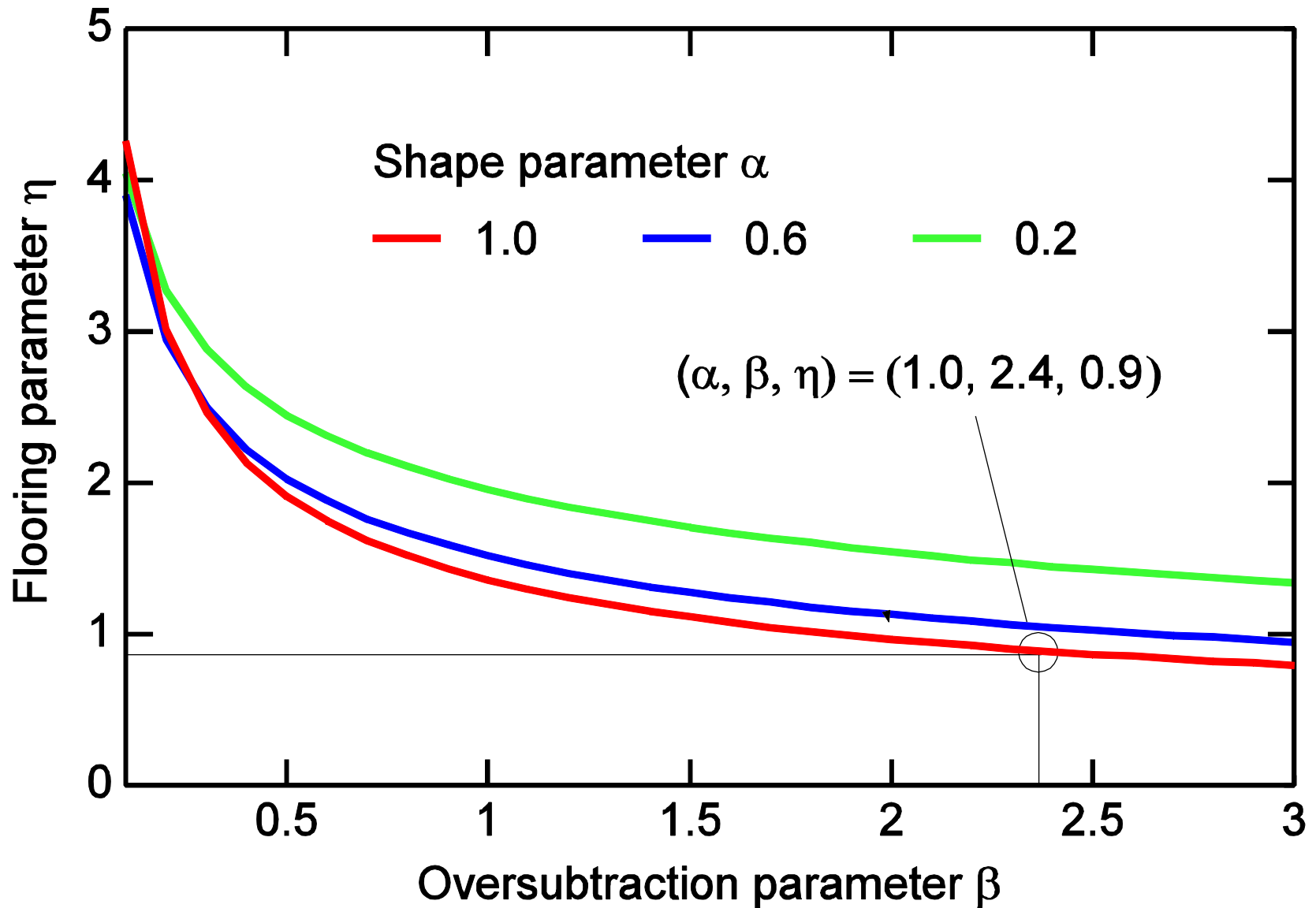
$$\left. \left\{ \mathcal{S}(\alpha_0, \beta, 4)(\alpha_0+1)\alpha_0 - \mathcal{S}^2(\alpha_0, \beta, 2)(\alpha_0+3)(\alpha_0+2) \right\}^{\frac{1}{2}} \right]$$

一部の信号処理

は代数的に最適化

を行うことができる

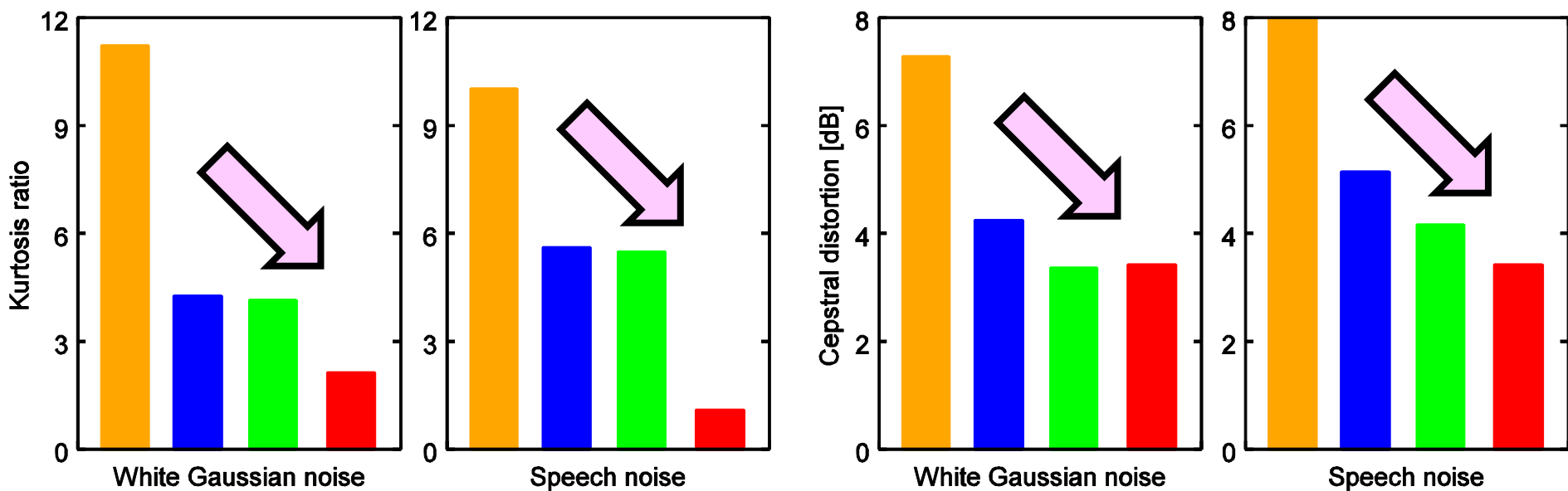
# ミュージカルノイズフリーを満たすパラメータの例



# 評価実験

目的話者	音素バランス文 4文 (男性2名, 女性2名)
雑音源	白色ガウス雑音 ( $\alpha = 0.97$ ) 話し声雑音 ( $\alpha = 0.21$ )
入力SN比	0 [dB]
サンプリング周波数	16 [kHz]
fft長	1024 [samples]
窓長	512 [samples]
シフト長	256 [samples]
NRR	10 [dB]

# 客観評価実験結果



従来の雑音抑圧手法 (one-shot SS, WF, MMSE-STSA) に比べて、ミュージカルノイズが少なく、音声歪みも少ない

# 主観評価実験結果

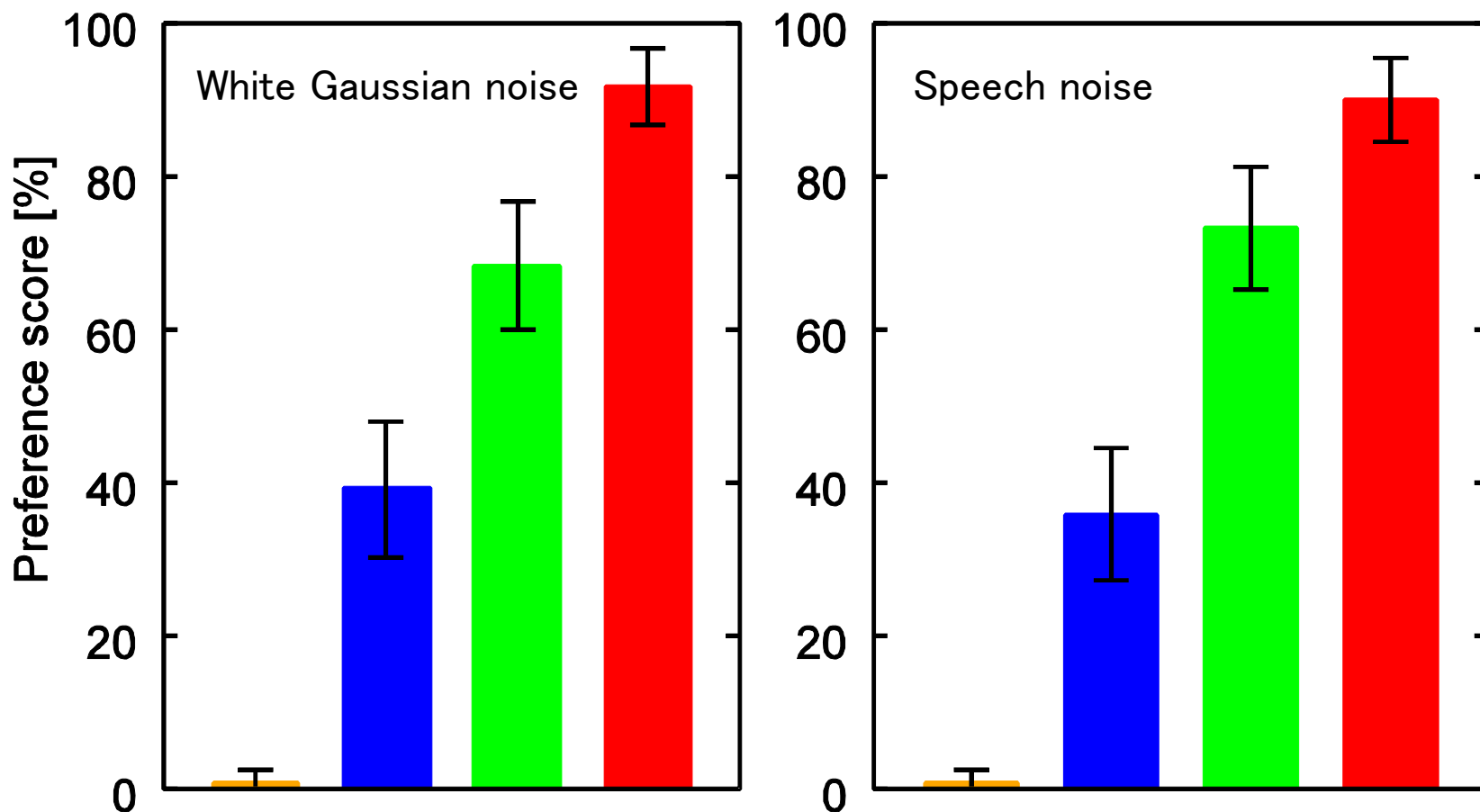
One-shot spectral subtraction: 

Wiener filtering: 

MMSE STSA estimator: 

Iterative spectral subtraction: 

95% confidence intervals: 



# 不動点発見におけるセレンディピティ

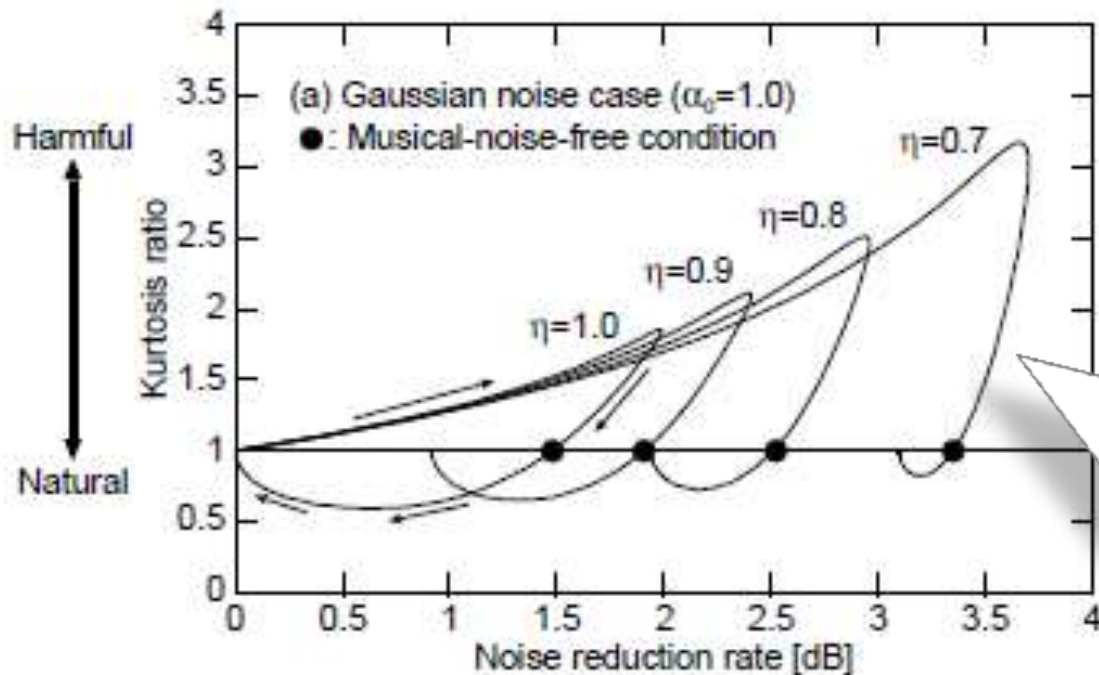
- 高次統計量(カートシス)の不動点はなぜ見つかったのか
  - 「不動点を見つけよう」と狙っていたわけではない
  - ある学生が「**雑音を過大に引き、かつ原音を過大に足すと、音色が元に戻る時がある!**」と報告してきた
  - 誰も信じなかった。学生のプログラムのバグだと思っていた
  - そもそも、雑音を減らすための信号処理において「雑音を足しなおす」というのは**常識的におかしい!**



# 不動点発見におけるセレンディピティ

## ■ 高次統計量空間における「ヒステリシス」の確認

- 本当にあっているかどうか数理的に確認してみた
- すると、なんと、高次統計量空間には非線形性に由来する**ヒステリシス現象が存在する**ことが分かった
- 存在が証明されれば後はそれが「何時生じるか」を予測するだけ



IEEE論文に投稿した際に、査読で「意味が分からなかったので削除せよ」と言われたヒステリシスの存在証明図。

後に国際会議招待講演にて喝采をあげた。