



東京大学 工学部 計数工学科/物理工学科

応用音響学：音声分析 (2) 短時間スペクトル

嵯峨山 茂樹 <sagayama@hil.t.u-tokyo.ac.jp>

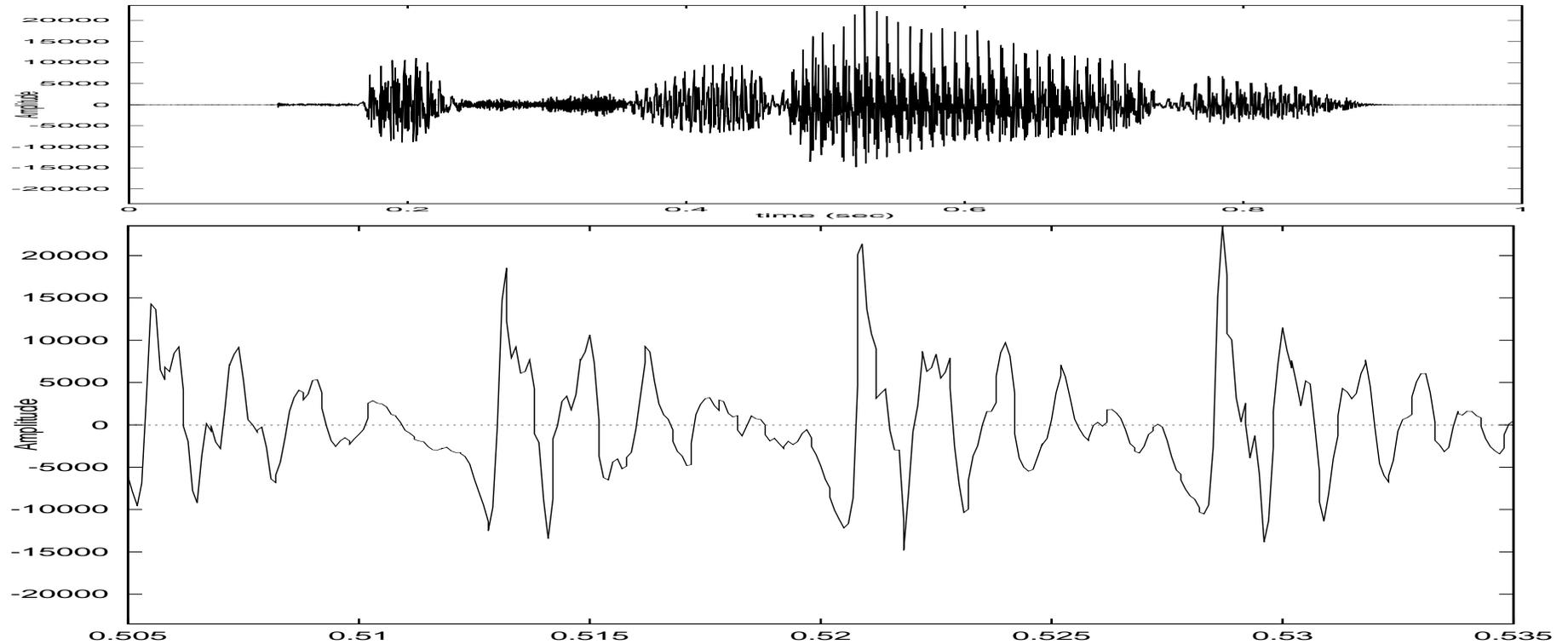
東京大学 工学部 計数工学科 <http://hil.t.u-tokyo.ac.jp/>

■ 要点

- データ窓
- 短時間自己相関関数
- 短時間パワースペクトル
- **Wiener-Khinchine** の定理と **Herglotz** の定理



音声の波形，高域強調，Hamming 窓の例



音声の波形，高域強調，Hamming 窓の例 (男声 /a/)

wavesurfer



音声信号の短時間スペクトル分析

■ 必要性

- スペクトル解析は、基本的に定常信号の理論。
- 音声などは非定常。非定常だから情報が運べる。
- 短区間を定常信号とみなした分析手法

■ 方法

- 高域強調 (フィルタ $1 - \alpha z^{-1}$)
- データ窓を掛ける 分析フレーム
(典型的条件: 30mS窓長、10mSシフト)
- 短区間に対して、自己相関関数、スペクトル解析



高域強調 (preemphasis)

■ 必要性

- 音声のスペクトルは全体的に傾斜。(母音の場合、約-6dB/oct)
- 音声の特徴はフォルマントで代表される。

■ 方法

- 高域強調フィルタ $1 - \alpha z^{-1}$
(通常、 $\alpha = 0.9 \sim 0.98$ 。嵯峨山 は $\alpha = 1 - 2^{-5} = 0.96875$)

(a) 男声「こしらえる」/koʃiraeru/ 中の /i/ の短時間スペクトル形状

(b) 高域強調した上記のスペクトル



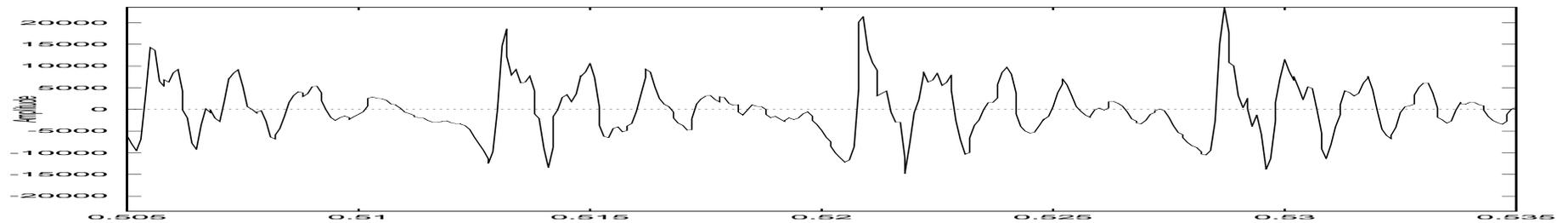
高域強調 (preemphasis) (続き)

■ 波形領域の処理

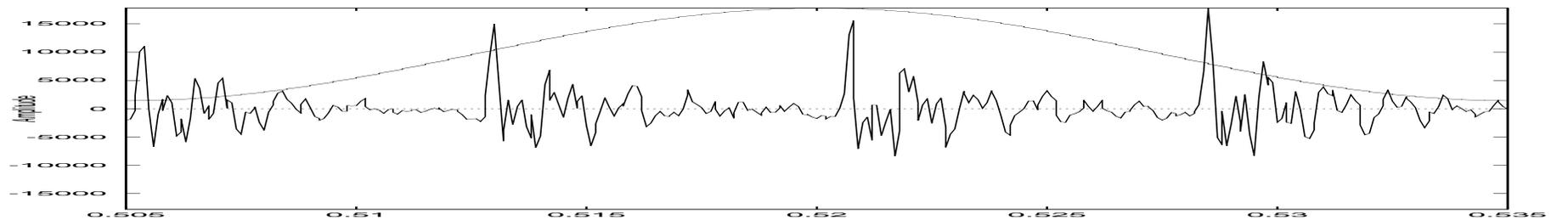
サンプル値列: $\{x_0, x_1, \dots, x_{N-1}\}$ に対して、

$$x'_t = x_t - \alpha x_{t-1}$$

(係数つき差分)



(c) 男声 第52フレーム /a/ 原波形



(d) 男声 第52フレーム /a/ 高域強調と Hamming 窓



高域強調 (preemphasis) (続き：効果の例)

(b) 300mS 付近 /f/

(c) 430mS 付近 /i/

(d) 520mS 付近 /a/

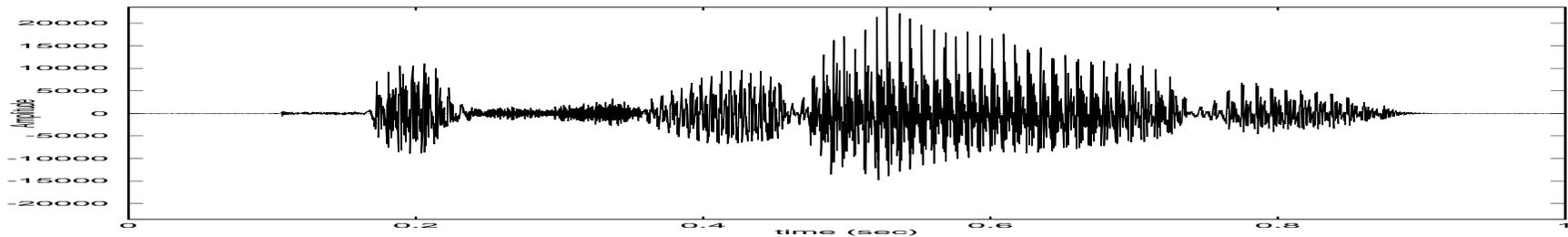
図：高域強調の例 – 特に有声音のスペクトルの平坦化の効果が見られる

(上段：高域強調なし、下段：高域強調あり (特性: $1 - 0.96875z^{-1}$))



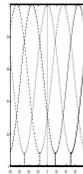
音声信号の短時間スペクトル分析: データ窓

- 短区間サンプル値列: $\{x_0, x_1, \dots, x_{N-1}\}$
- データ窓長 (30mS typ.)、データ窓シフト (10mS typ.)
- データ窓掛け: $\{y_0, y_1, \dots, y_{N-1}\}$, $y_t = w_t x_t$
 例: Hamming 窓 $w(\lambda) = 0.54 + 0.46 \cos \pi \lambda$, $|\lambda| \leq 1$

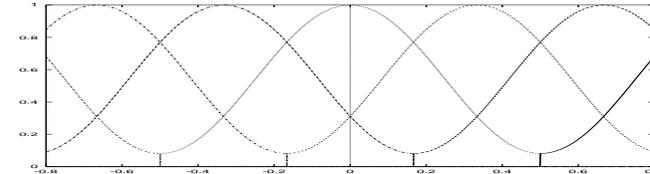


(a) 男声「こしらえる」/kojiraeru/ の信号波形

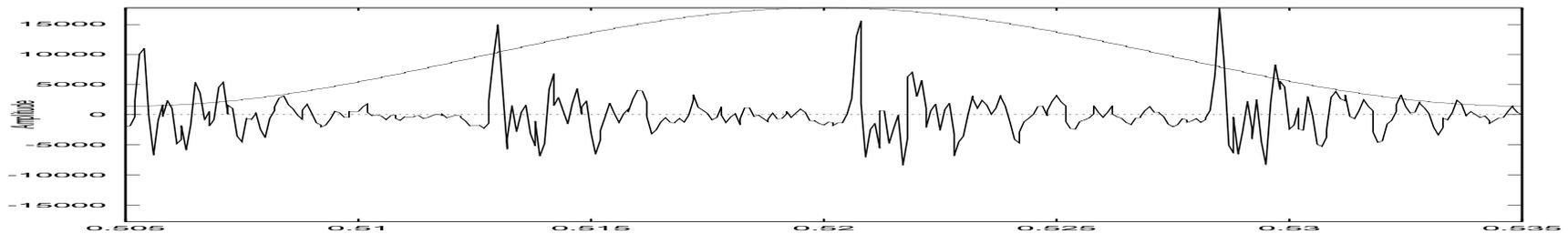
実際の窓幅:



拡大:



(b) フレームの相対関係。Hamming 窓を 1/3 シフトして適用する。横軸は窓長を 1 とした時刻。



(d) 男声 第 52 フレーム /a/ 高域強調と Hamming 窓



音声信号の短時間スペクトル (periodogram)

- 短時間標本スペクトル密度 (periodogram): $I_N(\omega)$

$$I_N(\omega) = \frac{1}{2\pi N} \left| \sum_{t=0}^{N-1} y_t e^{-j t \omega} \right|^2 : \text{正值対称関数}$$

- 計算法: FFT (Fast Fourier Transform) を用いて DFT (Discrete Fourier Transform) $S(\omega) = \sum_{t=0}^{N-1} y_t e^{-j t \omega}$ を計算し、絶対値二乗計算 $|S(\omega)|^2$ 。

(b) 300mS 付近 /s/

(c) 430mS 付近 /i/

(d) 520mS 付近 /a/

図: 短時間音声スペクトルの例 (高域強調あり)



音声信号の短時間自己相関関数

■ 短時間自己相関関数 (short-term autocorrelation function) $\{v_\tau\}$

$$v_\tau = \frac{1}{N} \sum_{t=\tau}^{N-1} y_t y_{t-\tau}, \quad v_{-\tau} = v_\tau, \quad \tau = 0, 1, 2, 3, \dots$$

■ 正定値性:

$$\sum \xi_i \xi_j v_{i-j} > 0 \quad \text{for any } \{\xi_i\}$$

「数列 $\{v_0, v_1, v_2, \dots\}$ が正定値列を成す。」

or Toeplitz型 行列

$$\mathbf{V} = \begin{pmatrix} v_0 & v_1 & \cdots & v_{p-1} \\ v_1 & v_0 & \cdots & v_{p-1} \\ \mathbf{i} & \mathbf{i} & \ddots & \mathbf{i} \\ v_{p-1} & v_{p-2} & \cdots & v_0 \end{pmatrix}$$

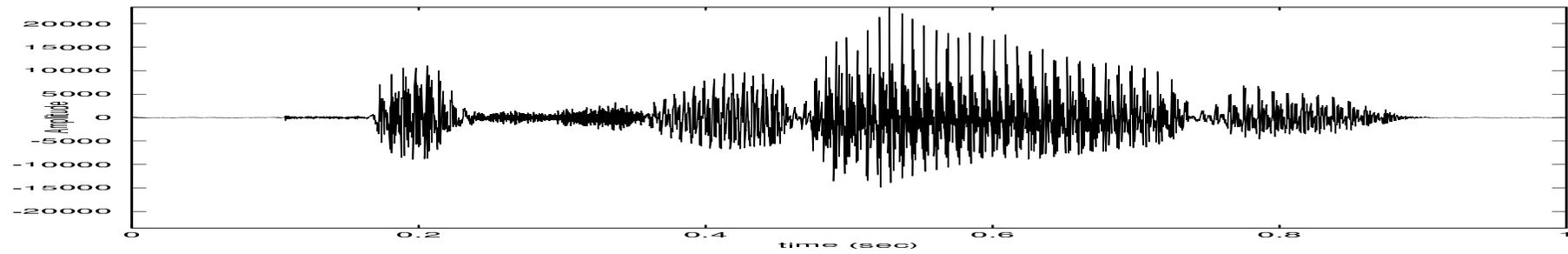
が正定値

■ Cf. 不偏自己相関関数推定: 正定値でない

$$v_\tau = \frac{1}{N - \tau} \sum_{t=\tau}^{N-1} y_t y_{t-\tau}$$



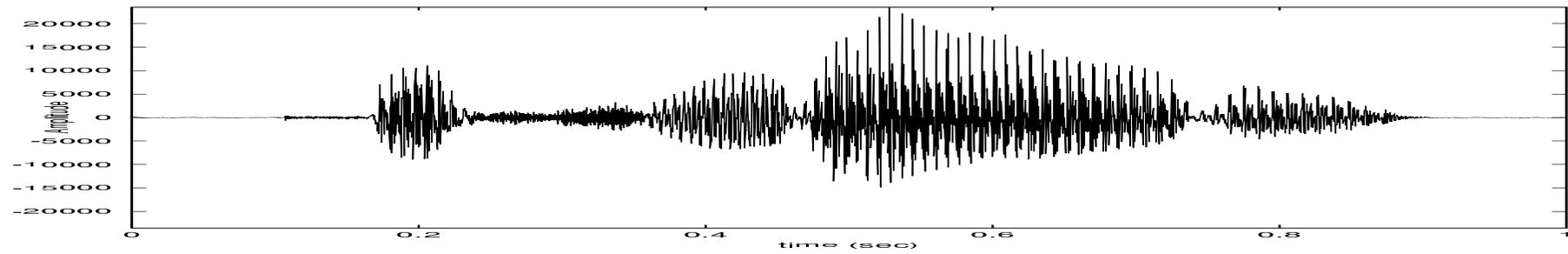
音声信号の短時間自己相関関数の例



図：音声の自己相関関数形状の例 (高域強調なし)。上から, (a) 音声波形 (男声 (MAU) 「こしらえる」), (b) 300mS 付近 /j/, (c) 430mS 付近 /i/, (d) 520mS 付近 /a/



高域強調音声信号の短時間自己相関関数の例



図：音声の自己相関関数形状の例 (高域強調あり)。上から, (a) 音声波形 (男声 (MAU) 「こしらえる」), (b) 300mS 付近 /j/, (c) 430mS 付近 /i/, (d) 520mS 付近 /a/



正定値自己相関関数とHerglotzの定理

1. 正定値短時間自己相関関数 $\{v_\tau\}$

$$v_\tau = \frac{1}{N} \sum_{t=\tau}^{N-1} y_t y_{t-\tau}, \quad v_{-\tau} = v_\tau, \quad \tau = 0, 1, 2, 3, \dots$$

Cf. 不偏推定 $v_\tau = \frac{1}{N - \tau} \sum_{t=\tau}^{N-1} y_t y_{t-\tau}, \quad v_{-\tau} = v_\tau, \quad \tau = 0, 1, 2, 3, \dots$

2. 正定値列 $\{v_\tau\}$: $\sum_{i,j} \xi_i \xi_j^* v_{i-j} > 0$ for any complex $\{\xi_i\}$ not all zero,

i.e., Toeplitz型 行列 $v = \begin{pmatrix} v_0 & v_1 & \cdots & v_{p-1} \\ v_1 & v_0 & \cdots & v_{p-1} \\ \vdots & \vdots & \ddots & \vdots \\ v_{p-1} & v_{p-2} & \cdots & v_0 \end{pmatrix}$ が正定値 ($p = 1, 2, 3, \dots$)

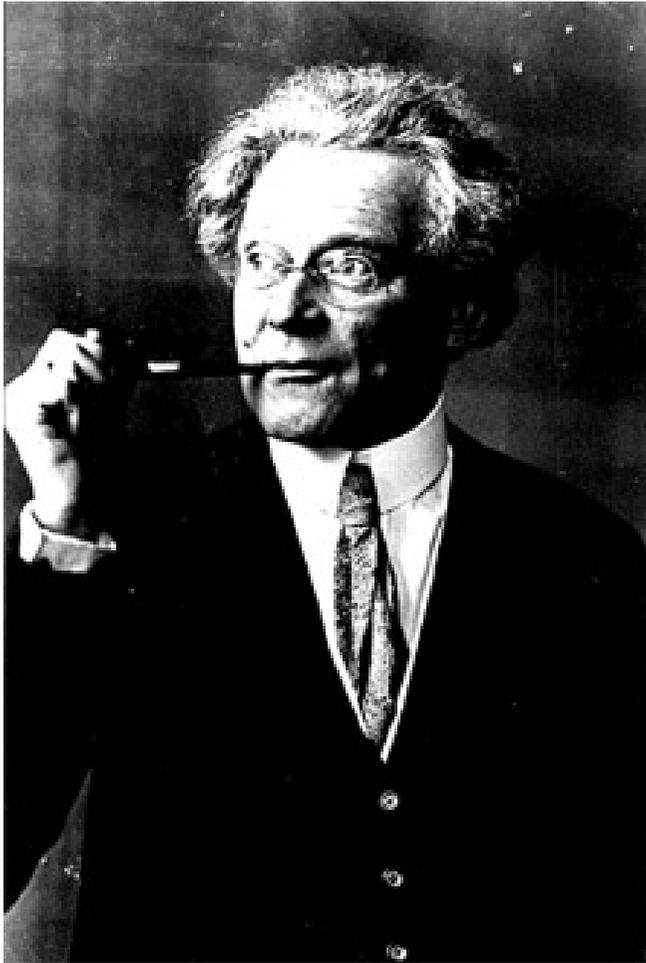
3. Herglotz の定理: 正定値列のFourier級数は正值対称関数

$\{v_\tau\}$ が正定値列 $\overset{\text{必要十分}}{\Leftrightarrow} I_N(\omega)$ が正值対称関数, **i.e.,**

$$I_N(\omega) = I_N(-\omega) > 0, \quad I_N(\omega) = \frac{1}{2\pi} \sum_{\tau=-N+1}^{N-1} v_\tau \cos \omega \tau$$



Herglotz



Gustav Herglotz

Gustav Herglotz wurde 1881 in Wallern (Böhmen) geboren. 1925 wurde er als Nachfolger von C. Runge nach Göttingen berufen. Er starb 1953 in Göttingen.



Periodogram と Wiener-Khinchine の定理

1. 短時間標本スペクトル密度 (periodogram): $I_N(\omega)$

$$I_N(\omega) = \frac{1}{2\pi N} \left| \sum_{t=0}^{N-1} y_t e^{-j t \omega} \right|^2 \quad \text{波形の DFT の } |\cdot| \text{ から計算}$$

$$= \frac{1}{2\pi} \sum_{\tau=-N+1}^{N-1} v_\tau \cos \omega, \quad \text{自己相関関数の } \cos \text{ 変換}$$

$$(-\pi < \omega < \pi)$$

2. Wiener-Khinchine の定理

「自己相関関数のフーリエ変換はパワースペクトル」

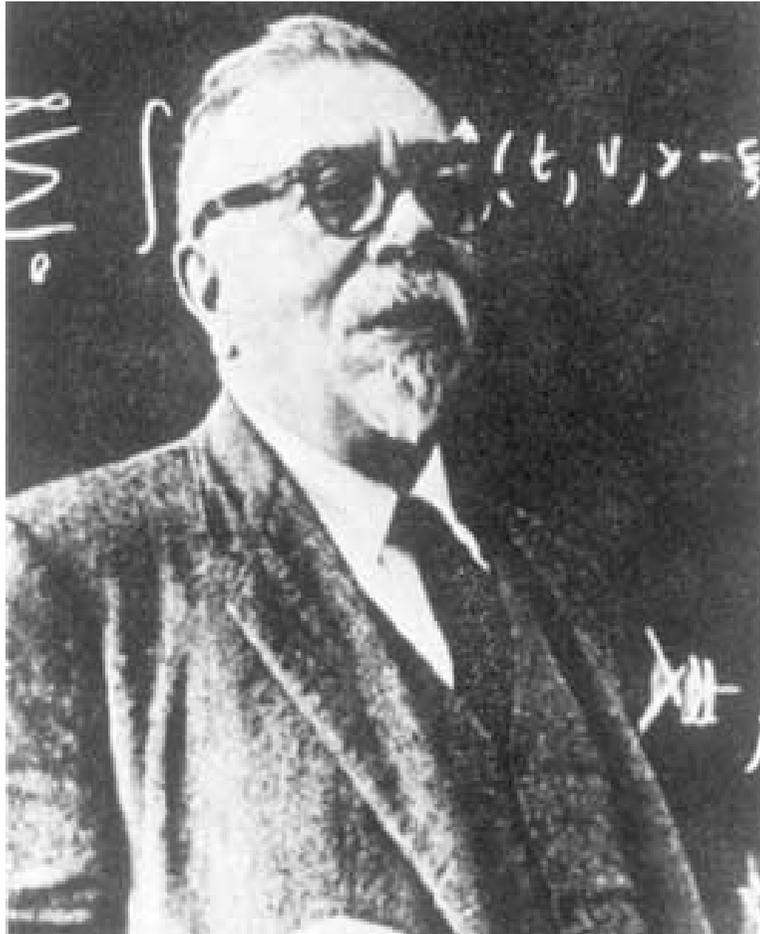
$$S(\omega) = \mathcal{F} [v(\tau)], \quad -\infty < \tau < \infty, \quad -\infty < \omega < \infty$$

離散時間では、

$$S(\omega) = \mathcal{F} [v_\tau], \quad \tau \in Z, \quad -\pi < \omega < \pi$$



Wiener と Khinchin



Norbert Wiener

Born: 26 Nov 1894 in Columbia, Missouri, USA

Died: 18 March 1964 in Stockholm, Sweden

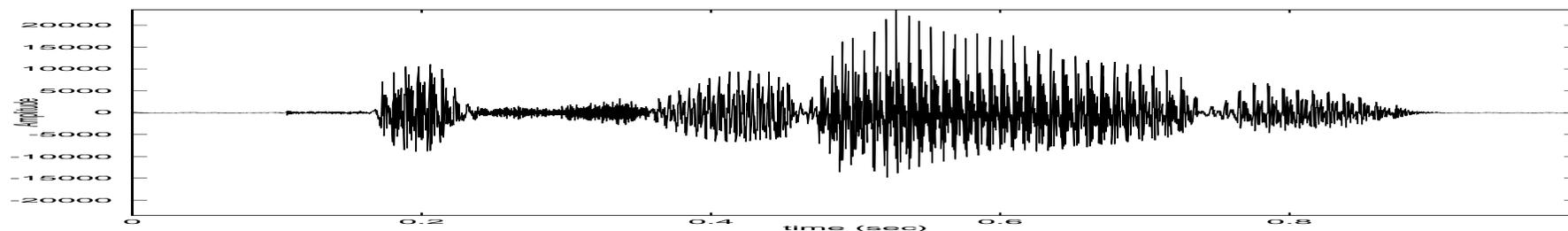


Aleksandr Yakovlevich Khinchin

Born: 19 July 1894 in Kondrovo, Kaluzhskaya guberniya, Russia

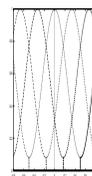
Died: 18 Nov 1959 in Moscow, USSR

音波形，高域強調，Hamming 窓，分析フレーム

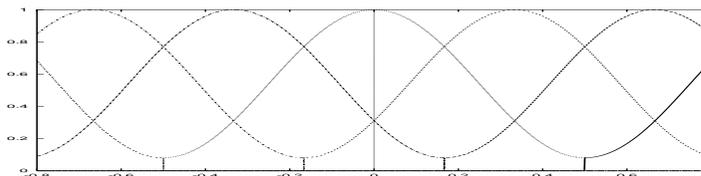


(a) 男声「こしらえる」/kɔʃiraeru/ の信号波形

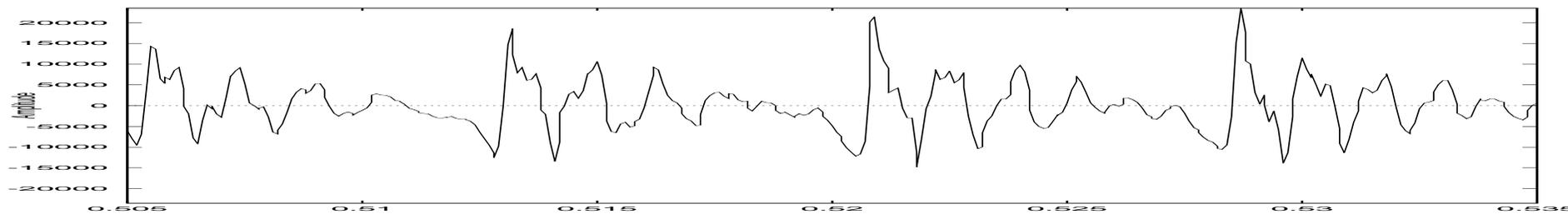
実際の窓幅:



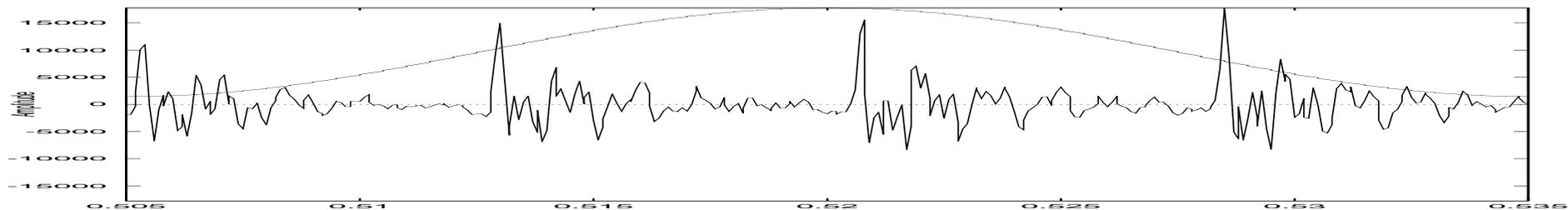
拡大:



(b) フレームの相対関係。Hamming 窓を 1/3 シフトして適用する。横軸は窓長を 1 とした時刻。



(c) 男声 第 52 フレーム /a/ 原波形



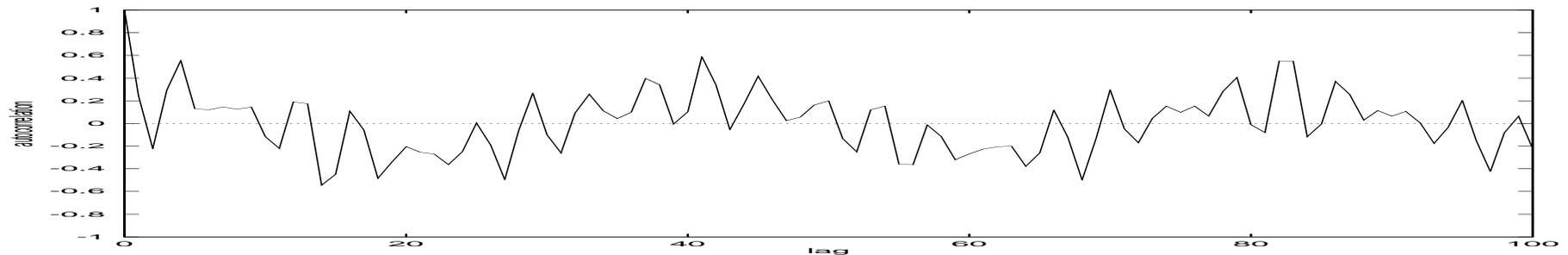
(d) 男声 第 52 フレーム /a/ 高域強調と Hamming 窓



音声信号の短時間スペクトルと自己相関関数

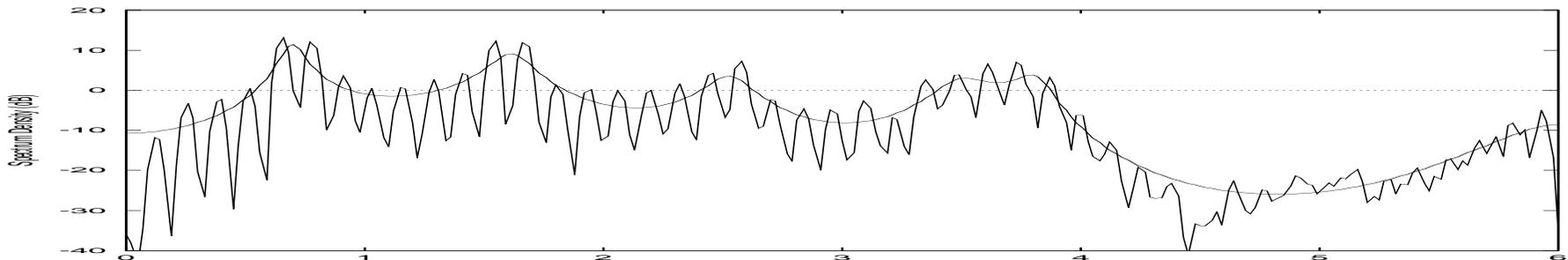
正定値標本自己相関関数 ($\sum \xi_i \xi_j v_{i-j} > 0$ for any $\{\xi_i\}$ or 行列 $v = \begin{pmatrix} v_0 & v_1 & \cdots & v_{p-1} \\ v_1 & v_0 & \cdots & v_{p-1} \\ \vdots & \vdots & \ddots & \vdots \\ v_{p-1} & v_{p-2} & \cdots & v_0 \end{pmatrix}$ 正定値)

$$v_\tau = \frac{1}{N} \sum_{t=\tau}^{N-1} y_t y_{t+\tau}, \quad v_{-\tau} = v_\tau, \quad \tau = 0, 1, 2, 3, \dots$$



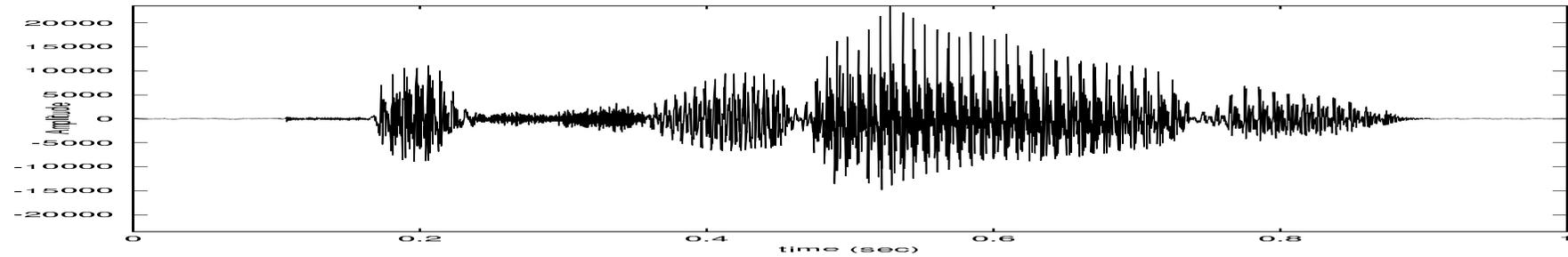
短時間標本スペクトル密度 (periodogram) (正値対称関数 $f(\omega) = f(-\omega) > 0$)

$$I_N(\omega) = \frac{1}{2\pi N} \left| \sum_{t=0}^{N-1} y_t e^{-j t \omega} \right|^2 = \frac{1}{2\pi} \sum_{\tau=-N+1}^{N-1} v_\tau \omega, \quad -\pi < \omega < \pi$$





音声信号の短時間スペクトルの例



(a) 音声波形(男声(MAU)「こしらえる」)

(b) 300mS 付近 /j/

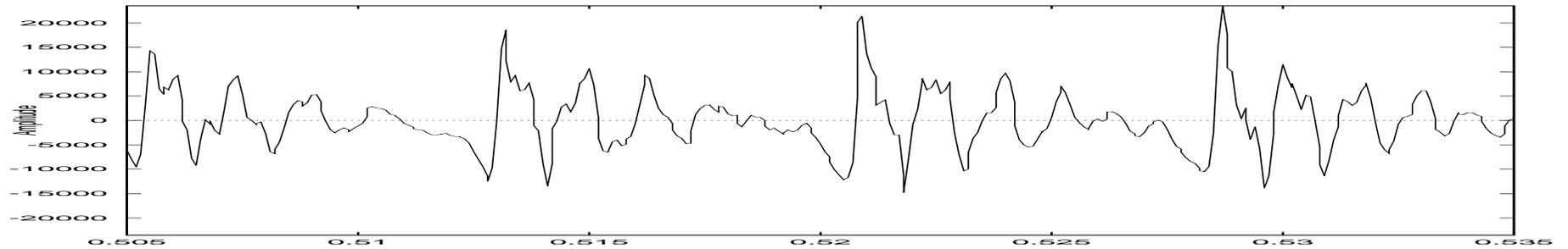
(c) 430mS 付近 /i/

(d) 520mS 付近 /a/

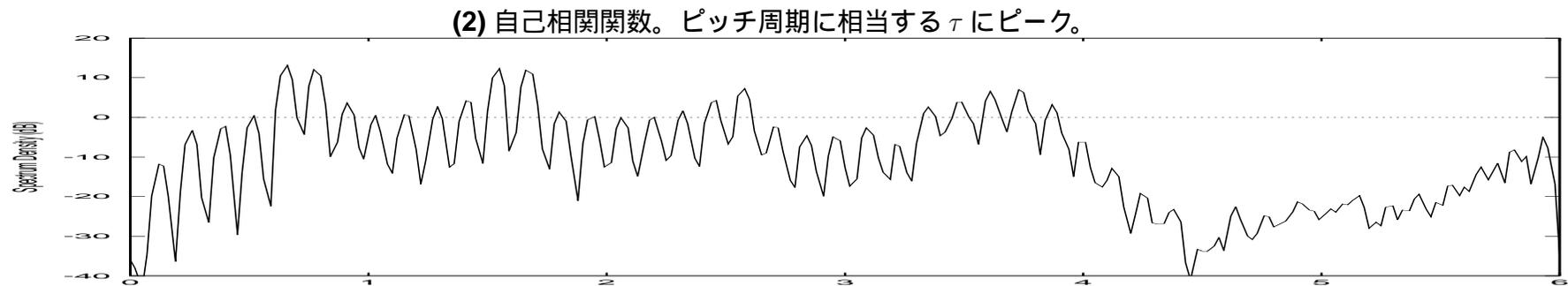
図: 音声の短時間スペクトルの例 (高域強調なし)



音声の短時間波形、自己相関関数、スペクトルの例



(1) 音声 (男声 /a/) の短時間波形。ほぼ周期的な波形。(高域強調なし)



(2) 自己相関関数。ピッチ周期に相当する τ にピーク。

(3) スペクトル (periodogram) を対数表示。ピッチに相当する細かい山谷と、共振特性の山 (フォルマント) が見られる。

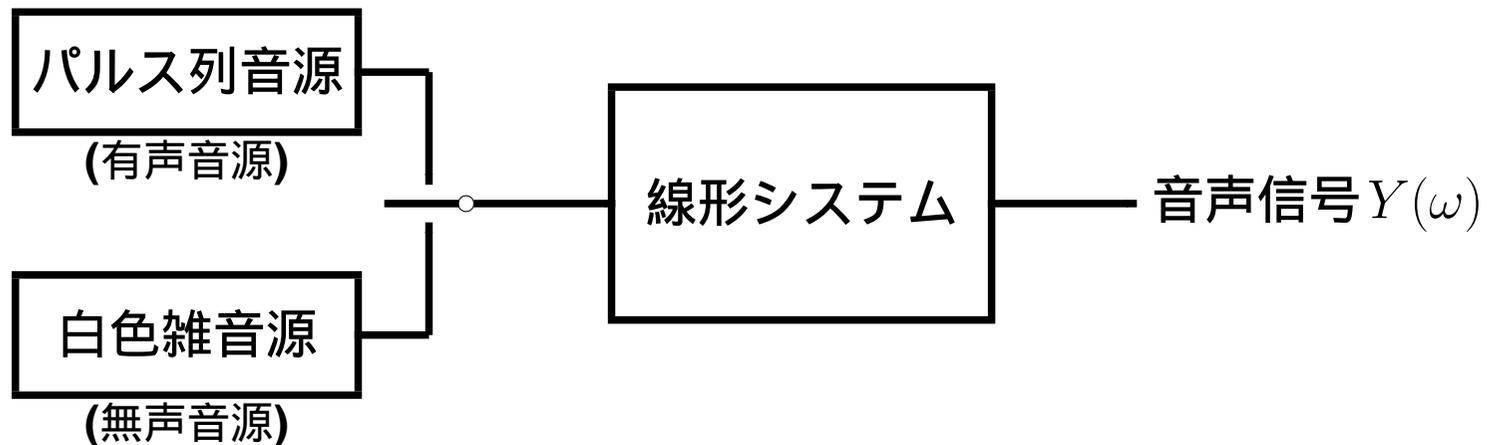


音声生成の線形モデル

- 有声音: $\left\{ \begin{array}{l} \text{母音 [a, i, u, e, o, (\text{æ, ɛ, ɔ, ɒ, ʌ, ɜ, \text{ɛ, ɪ, y, \text{œ, ʏ, ʊ, U, ɔ, \text{ə, ...)] ,} \\ \text{有声音子音 [g, z, d, b, n, m, r, ʃ, (l, v, \text{ð, ʎ, ʌ, R, ɹ, ...)] ,} \\ \text{半母音 [w, j, (\text{ɥ})] } \end{array} \right.$
- 無声音: 無声音子音 [k, s, t, h, \text{ç, ʦ, \text{p, ç, \text{φ, (ʃ, f, \text{θ, β, ...)] }
- 線形系モデル

駆動音源部 $S(\omega)$

声道共振部 $H(\omega)$



- 波形領域、自己相関領域、スペクトル領域

$$y(t) = s(t) * h(t), \quad \phi_y(\tau) = \phi_s(\tau) * \phi_h(\tau), \quad Y(\omega) = S(\omega)H(\omega)$$