

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 (3) ケプストラム

嵯峨山 茂樹 <sagayama@hil.t.u-tokyo.ac.jp>

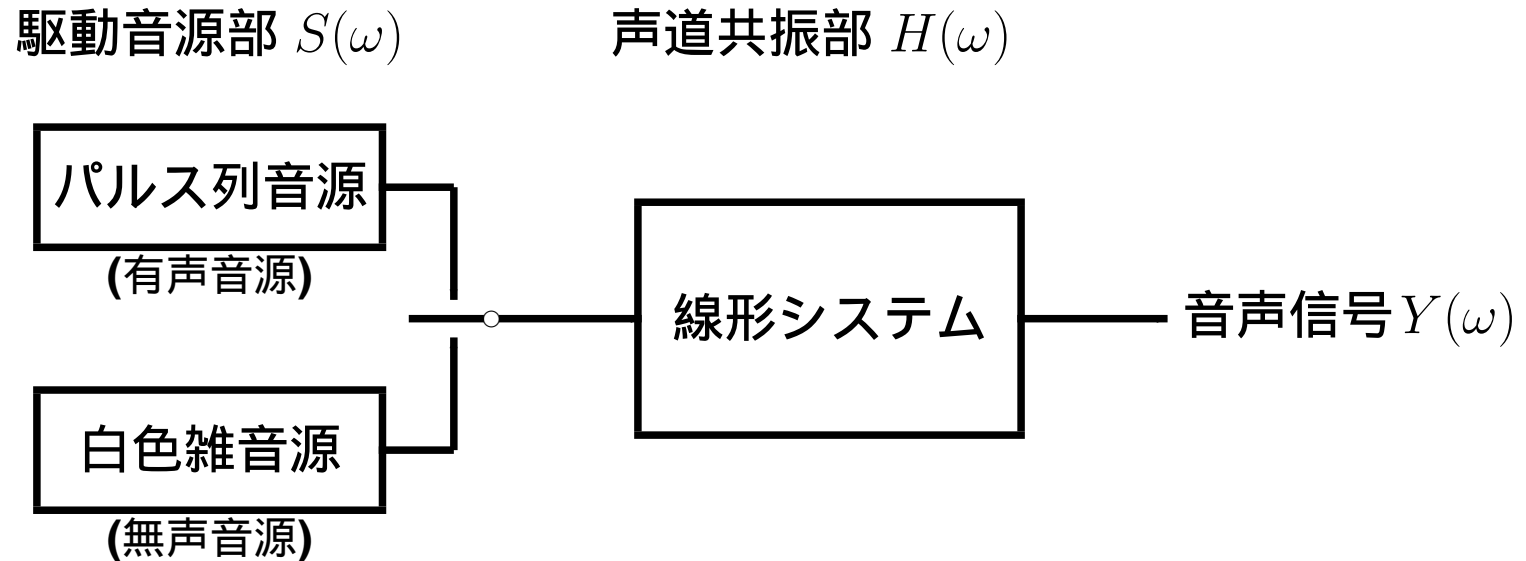
東京大学 工学部 計数工学科 <http://hil.t.u-tokyo.ac.jp/>

■ 要点

- ケプストラム
- 音声の短時間ケプストラム分析
- ケプストラムベクトル

音声生成の線形モデル

■ 線形系モデル



波形領域 $y(t) = s(t) * h(t)$

■ 自己相関領域 $\phi_y(\tau) = \phi_s(\tau) * \phi_h(\tau)$

スペクトル領域 $Y(\omega) = S(\omega)H(\omega)$

対数スペクトル

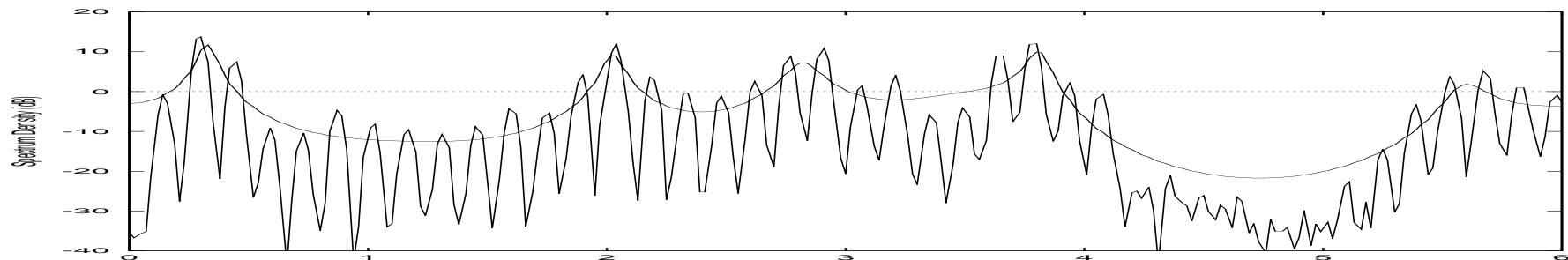
■ 線形スペクトル $f(\omega)$

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

■ 対数スペクトル $\log f(\omega)$

$$\log Y(\omega) = \log S(\omega) + \log H(\omega)$$

■ ピッチ構造と包絡形状



(a) 男声「こしらえる」/kojiraeru/ 中の /i/ の短時間スペクトル形状 (高域強調あり) と LPC により求めたスペクトル包絡

ケプストラム (cepstrum)

■ 自己相関関数: 線形スペクトル(spectrum)の逆フーリエ変換

$$v_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{jn\omega} f(\omega) d\omega \quad n = 0, 1, \dots$$

$$f(\omega) = \sum_{n=-\infty}^{\infty} v_n e^{-jn\omega} \quad -\pi \leq \omega \leq \pi$$

(Wiener-Khinchine の定理)

■ ケプストラム: 対数スペクトル(spectrum)の逆フーリエ変換

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{jn\omega} \log f(\omega) d\omega \quad n = 0, 1, \dots$$

$$\log f(\omega) = \sum_{n=-\infty}^{\infty} c_n e^{-jn\omega} \quad -\pi \leq \omega \leq \pi$$

■ Tukey ら (1963)

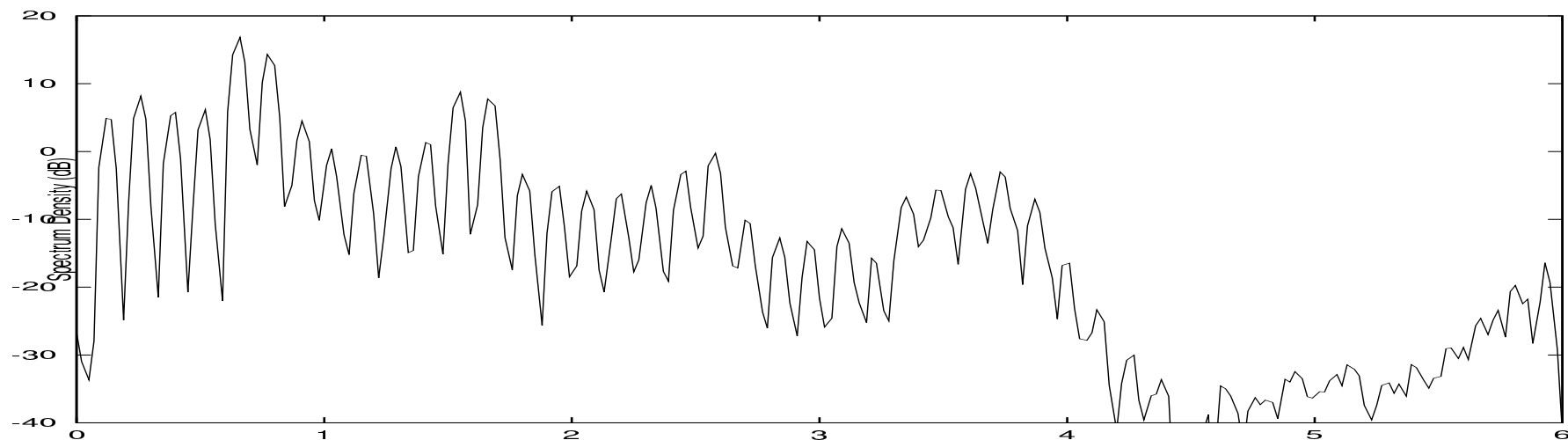
B. P. Bogert, M. J. R. Healy, and J. W. Tukey, "The Quefreny Alanysis of Time Series for Echoes: Cepstrum, Pseudo-autocovariance, CrossCepstrum, and Saphe Cracking," In M. Rosenblatt, editor, Proc. Syrup. Time Series Analysis, pages 209-243, New York, 1963. John Wiley and Sons.

Cepstrum 用語

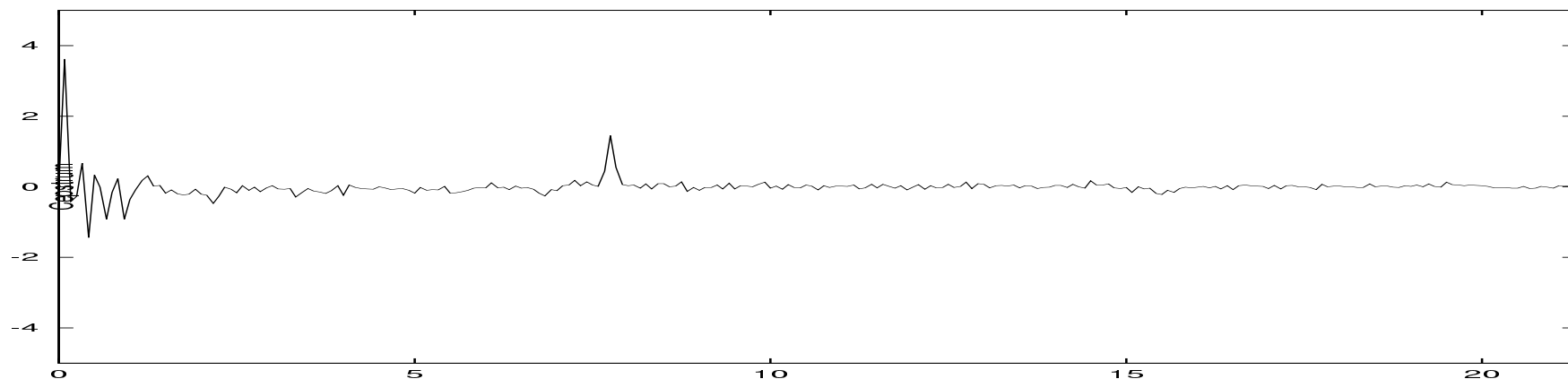
■ スペクトルとケプストラムの用語対比

名称	スペクトル (spectrum)	ケプストラム (cep strum)
横軸	周波数 (frequency)	ケフレンシ (que frequency)
操作	フィルタ (filter)	リフタ (lifter)
分析	アナリシス (analysis)	アラニシス (alan ysis)

音声信号の短時間ケプストラムの例



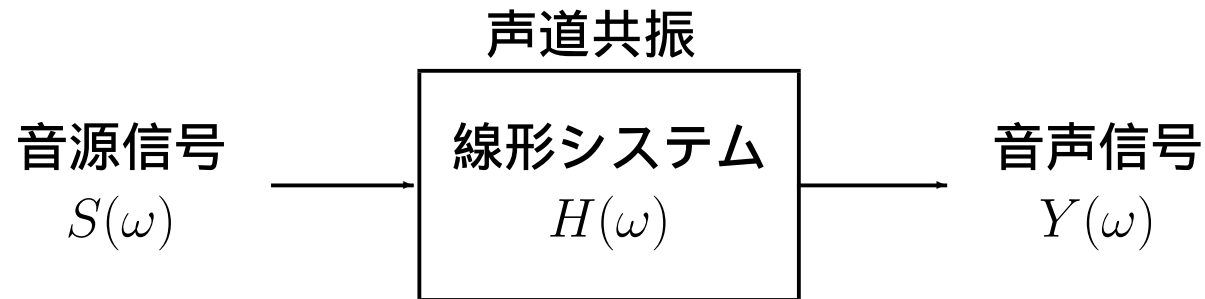
(a) 音声短時間スペクトル /a/



(b) 音声短時間ケプストラム /a/

図: 音声の短時間スペクトルとケプストラムの例 (男声 (MAU) 「こしらえる」 高域強調なし、520mS 付近)

線形系の縦続に関するケプストラムの性質



■ 線形系の接続

スペクトル(spectrum)領域: $S(\omega)H(\omega) = Y(\omega)$

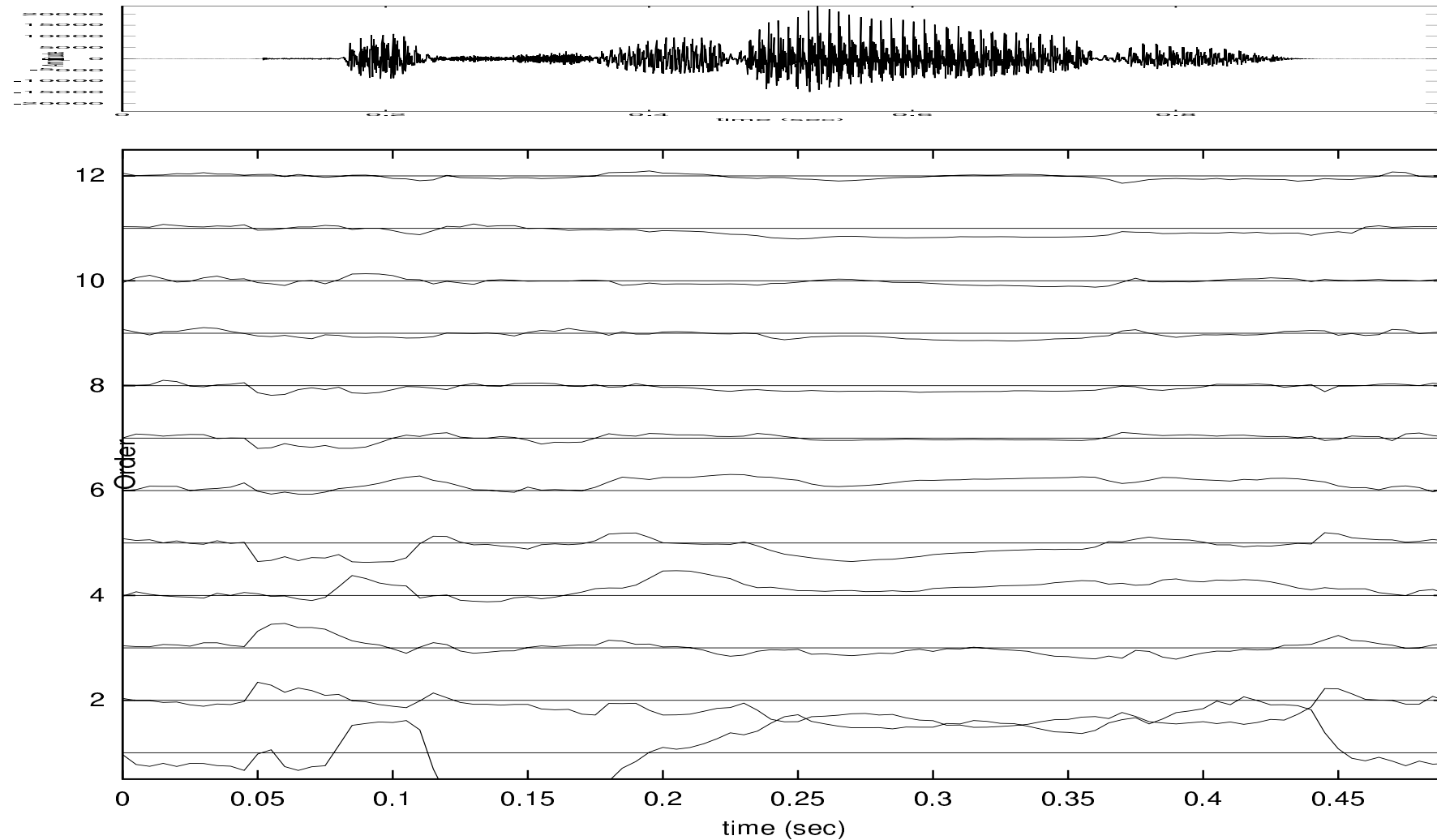
対数スペクトル領域: $\log |S(\omega)| + \log |H(\omega)| = \log |Y(\omega)|$

ケプストラム(cepstrum)領域: $c_n^S + c_n^H = c_n^Y, \quad n = 0, 1, 2, \dots$

■ ケプストラムの特徴: 線形系の従属接続のケプストラムは, それぞれのケプストラムの和

$$c_n^A + c_n^B = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{jn\omega} \log(f^A(\omega)f^B(\omega))d\omega \quad n = 0, 1, \dots$$

音声のケプストラム分析例



音声波形 (「こしらえる」) とケプストラム分析例 - 横軸は時刻、ケプストラムの次元毎に時間変換をプロット。

ケプストラム距離の性質

- Parseval の公式: ケプストラムの2乗和は, 対数スペクトルの2乗積分値に等しい:

$$\sum_{n=-\infty}^{\infty} |c_n|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \{\log f(\omega)\}^2 d\omega$$

- ケプストラムベクトル間のユークリッド距離

$$\sum_{n=-\infty}^{\infty} |c_n^A - c_n^B|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \{\log f^A(\omega) - \log f^B(\omega)\}^2 d\omega$$

対数スペクトル空間のユークリッド距離は, ケプストラム空間のユークリッド距離

- 有限次元ケプストラムベクトル間のユークリッド距離
 スペクトル概形間の2乗積分距離
 現代の音声認識の基礎
 さらに、周波数軸を聴覚特性に合わせて伸縮: “mel-cepstrum”

参考: スペクマート (specmurt)

- ケプストラム: 対数スペクトル(spectrum)の線形周波数での逆フーリエ変換

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{jn\omega} \log f(\omega) d\omega \quad n = 0, 1, \dots$$

$$\log f(\omega) = \sum_{n=-\infty}^{\infty} c_n e^{-jn\omega} \quad -\pi \leq \omega \leq \pi$$

- スペクマート: 線形スペクトル(spectrum)の対数周波数での逆フーリエ変換(式は少し怪しい)

$$y_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{jn\omega} f(\omega) d\log\omega \quad n = 0, 1, \dots$$

$$f(\omega) = \sum_{n=-\infty}^{\infty} y_n e^{-jn\log\omega} \quad -\pi \leq \omega \leq \pi$$

- 2003年 嵯峨山ら

高橋 佳吾, 西本 卓也, 嵯峨山 茂樹: “対数周波数逆畳み込みによる多重音の基本周波数解析,” 情報処理学会研究報告, 2003-MUS-53, pp.61-66, Dec. 2003.

スペクマート (specmurt) 用語

■ スペクトル、ケプストラム、スペクマートの用語対比

名称	スペクトラム (spectrum)	ケプストラム (cepstrum)	スペクマート (specmurt)
横軸	周波数 (frequency)	ケフレンシ (quefrequency)	フリンシク (frenyque)
縦軸	線形量	対数量	線形量
操作	フィルタ (filter)	リフタ (lifter)	フィルレット (filret)
分析	アナリシス (analysis)	アラニシス (alanysis)	アナシリス (anasylis)
創始	—	Tukey ら 1963	嵯峨山ら 2003

スペクマート (specmurt) 例

J. S. Bach: Ricercare à 6 from Musikalisches Opfer, BWV 1079, から(演奏)



図1. 該当する楽譜部分

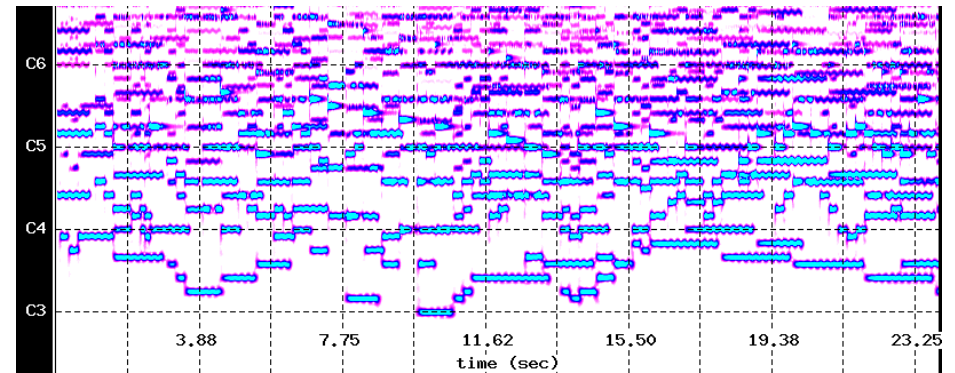


図3. 原スペクトル(wavelet分析)

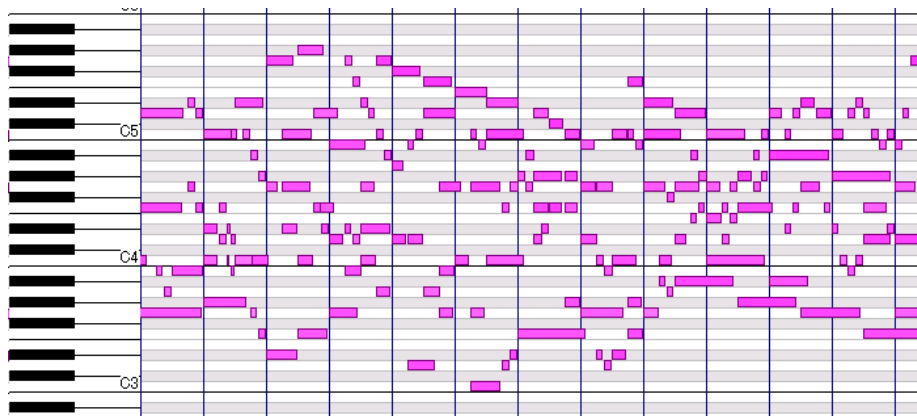


図2. Piano-roll表示 (手作業で作成)

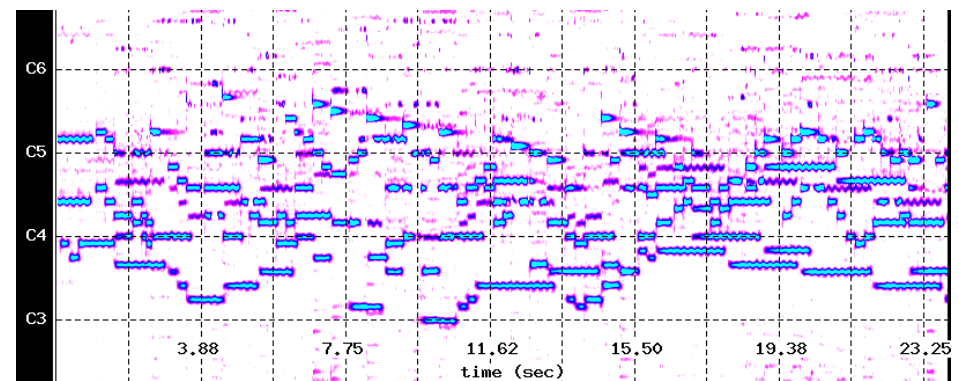


図4. specmurt分析

動的特徴パラメータ (1979)

- 音声パターン: 音声スペクトルパラメータのベクトル軌跡 接線ベクトルは動的特徴
- 音声信号の知覚: 瞬時ごとのスペクトルと、スペクトルの変化
- 動的特徴パラメータ: ベクトル時系列に部分的に重み窓つき最小自乗近似直線の傾き (回帰係数)

(嵯峨山、板倉「動的尺度に含まれる個人性情報」日本音響学会講演論文集 1979年6月。)

直線モデル: $\hat{\theta}(t) = a t + b + \text{誤差}$

対称重み $\{w_t\}$ の場合, 最小二乗法により

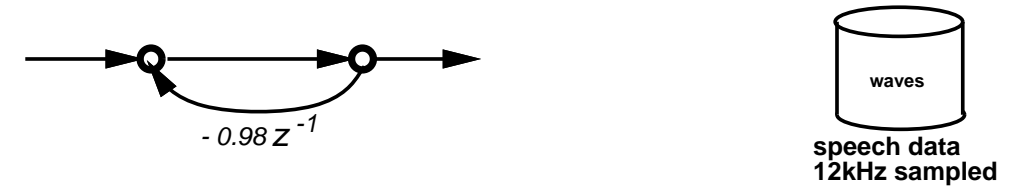
$$a = \frac{\sum_{t=-T}^T t w_t \theta(t)}{\sum_{t=-T}^T t^2 w_t}, \quad b = \frac{\sum_{t=-T}^T w_t \theta(t)}{\sum_{t=-T}^T w_t}$$

- a : $\theta(t)$ がケプストラムの場合: 「デルタケプストラム」 (音声認識誤り率が半減)

音声認識の標準的音声分析フロントエンド

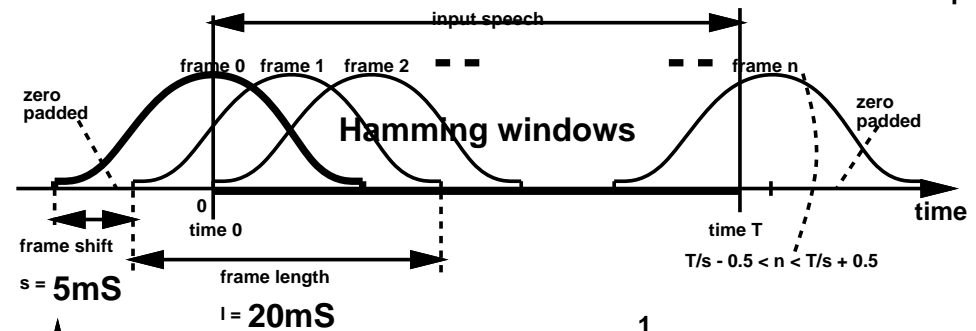
1 - Preemphasis

- Preemphasized with a high emphasis filter: $1 - 0.98z^{-1}$



2 - Time windowing

- 20 mS length Hamming window with 5 mS shift
- Started with frame 0 and ended with a frame the center of which is closest to the end point

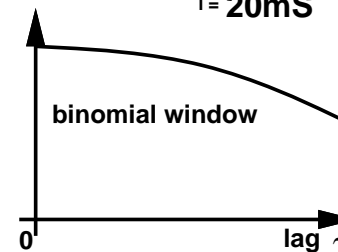


3 - Autocorrelation

- Positive definite autocorrelation, 20 points

4 - Lag windowing

- Binomial window (approximating Gaussian)
- Half band width = $0.01 * \text{sampling frequency}$, i.e., 120 Hz



5 - LPC analysis and conversion to LPC cepstra

- Solving a normal equation (correlation method) using the PARCOR recursive algorithm, $p = 16$.
- Conversion into LPC cepstrum by Oppenheim's recursive algorithm, $n = 16$.

6 - Weighted regression analysis of the cepstrum vector sequence

- Delta-cepstrum is derived from an LPC cepstrum vector sequence
- Regression time window is a 50 mS-long triangular window
- Weighted sum of squared fitting errors is minimized

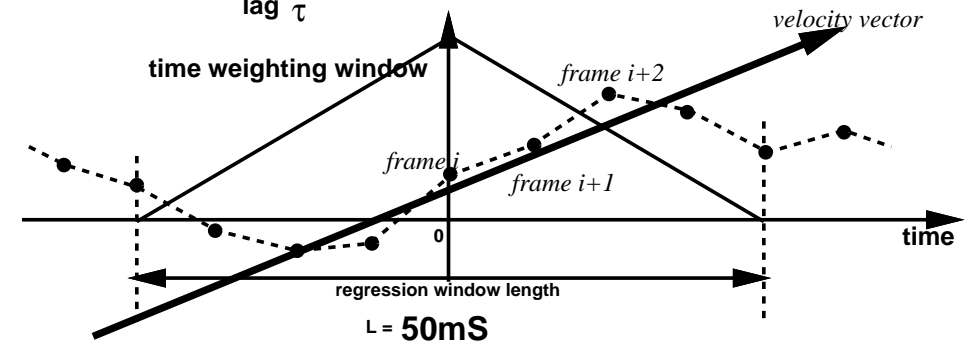


図5. 音声分析手順例

音声認識の標準的音声分析フロントエンド

音声入力

AD変換

アナログ音声波形をデジタル化

高域強調

スペクトルを平坦化 ($1 - \alpha z^{-1}$)

時間窓かけ

Hamming 窓 , Kaiser 窓

自己相関関数計算

積和計算

ラグ窓かけ

二項係数窓

LPC分析

Levinson-Durbin-板倉 アルゴリズム

ケプストラム変換

Oppenheim の漸化式

Δ ケプストラム計算

回帰係数計算

音声認識部へ