

「聴覚脳プロジェクト」における ウェーブレットの応用と展開

河原英紀

和歌山大学 / ATR / CREST

聴覚脳プロジェクト

n 科学技術振興事業団
戦略的基礎研究推進事業（CREST）
『脳を創る』研究領域
1997年度採択プロジェクト

聴覚の情景分析に基づく
音響・音声処理システム

聴覚脳プロジェクト

n 科学技術振興事業団
戦略的基礎研究推進事業（CREST）
『脳を創る』研究領域
1997年度採択プロジェクト

聴覚脳プロジェクト

プロジェクトの目標

n 人間の聴覚系と同型の処理 / 情報表現に基づく音響信号加工システムの実現

—音声・音響知覚研究のツール

n 高度に非線形なシステム
生態学的に妥当な刺激

—究極の聴覚補綴技術

—コンテンツ制作

n 聴覚の計算理論の形成

発表の概要

n デモ：何ができるようになるか？

n 聴覚に本質的な量をどう表現し求めるか？

—音色の知覚における不変性

n stabilized wavelet-Mellin transform and gammachirp

—ピッチ知覚と基本周波数

n instantaneous frequency of wavelet analysis

—音源の知覚と音響イベント

n multi-resolution zero-crossing and causality



原音声

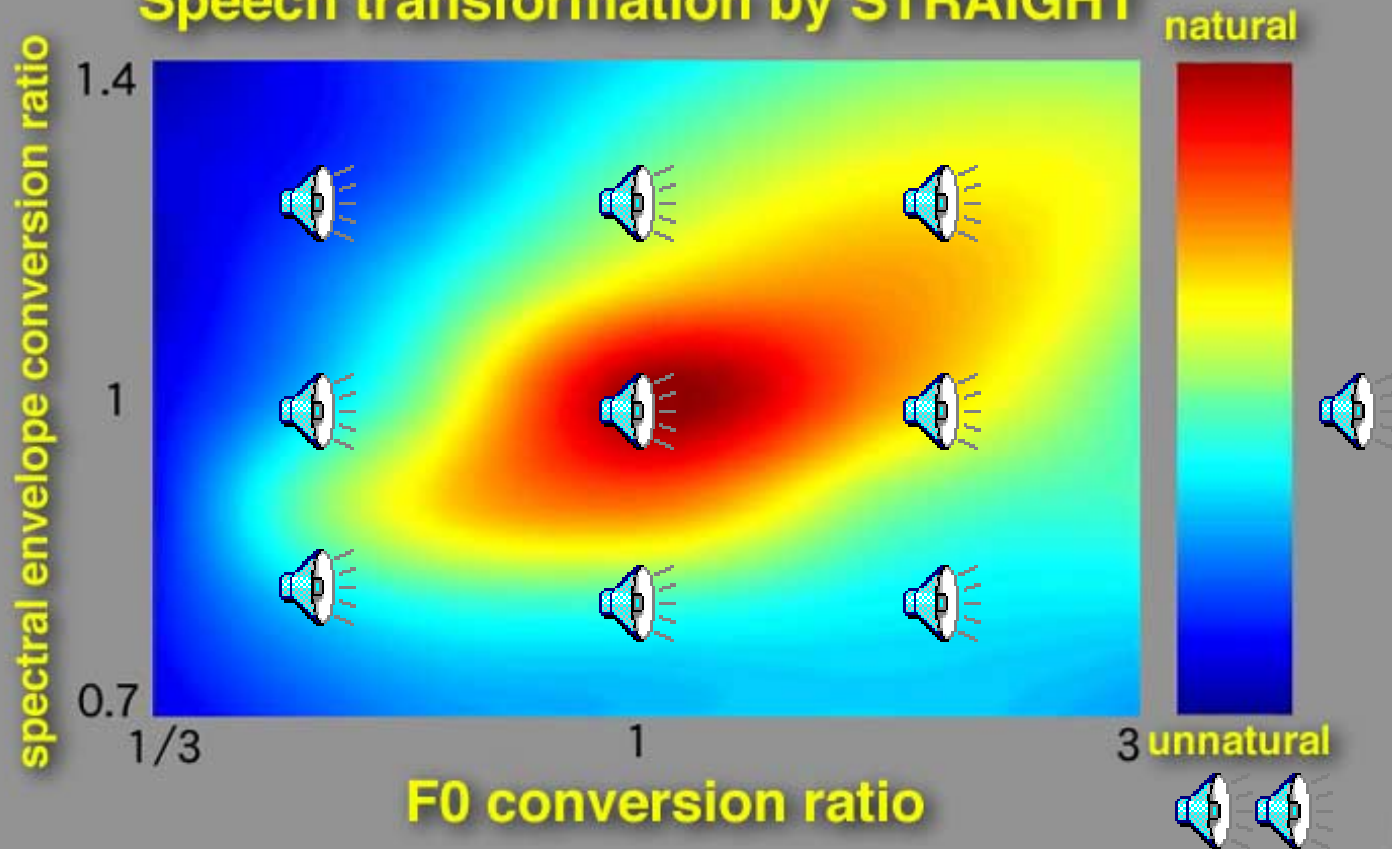
何が出来るようになるか？：現状

小

身体
の
大き
さ

大

Speech transformation by STRAIGHT



声帯の振動周波数

発表の概要

n デモ：何ができるようになるか？

n 聴覚に本質的な量をどう表現し求めるか？

—音色の知覚における不変性

n stabilized wavelet-Mellin transform and gammachirp

—ピッチ知覚と基本周波数

n instantaneous frequency of wavelet analysis

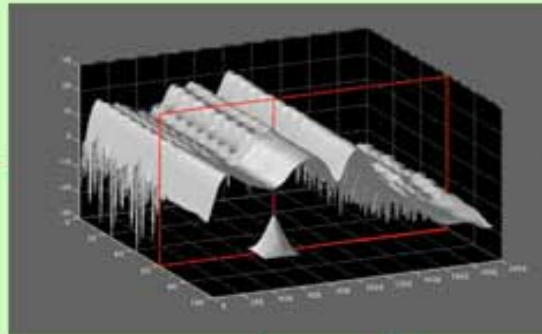
—音源の知覚と音響イベント

n multi-resolution zero-crossing and causality

STRAIGHT

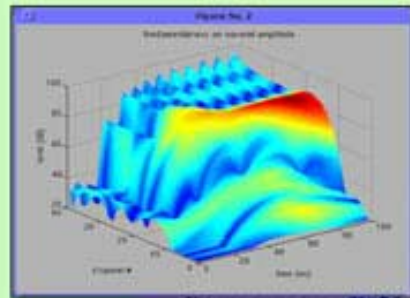
入力信号

F0-adaptive time-frequency smoothing
to eliminate periodicity interferences



出力音声

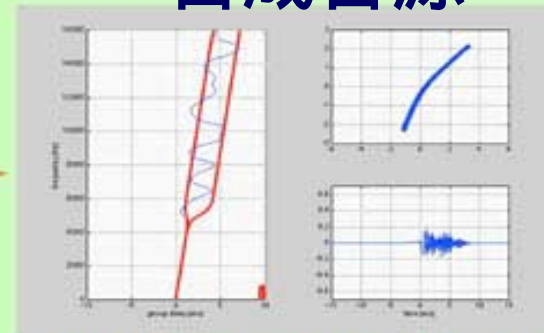
伝達特性



音源情報：
周期性等

Instantaneous-frequency-based
F0 and source information extractor

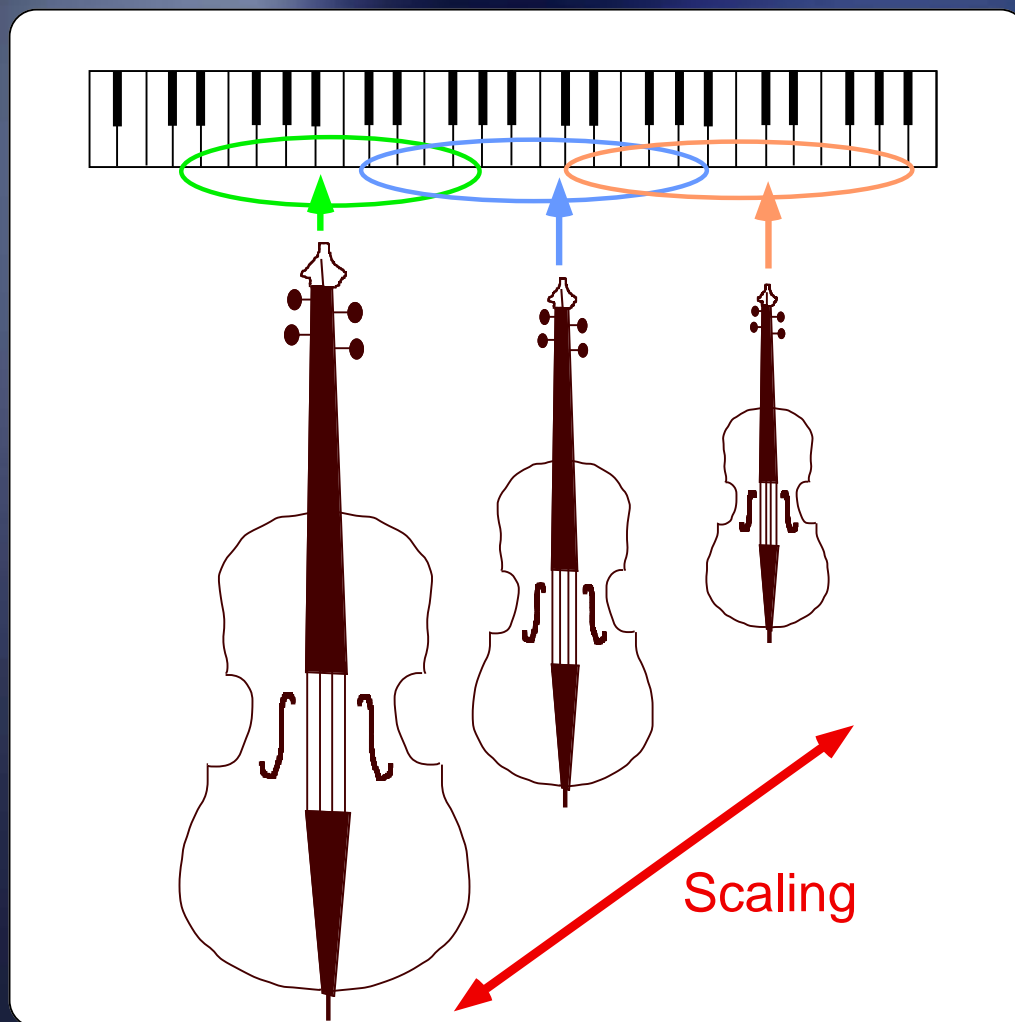
合成音源



Group delay manipulation to add
artificial naturalness

STRAIGHT is a very high-quality VOCODER.

楽器の種類・寸法と音色



類似の形状と構造
寸法の違い



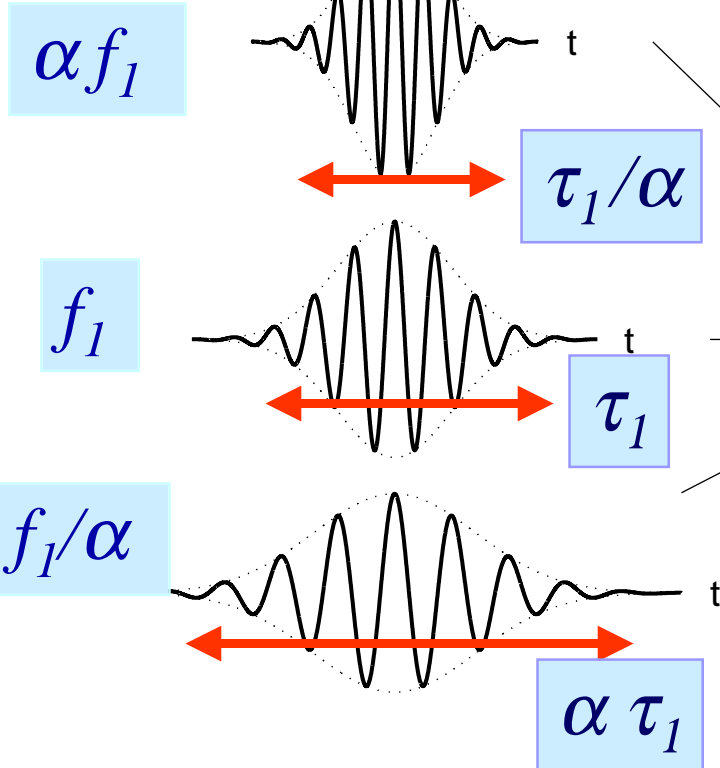
類似した音色

寸法変化に不変な性質を抽出する変換としての Mellin 変換

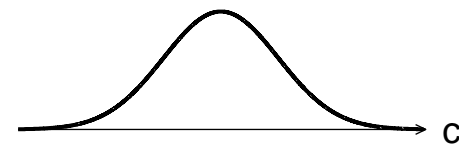
wavelet

Signal

継続時間 τ と搬送周波数 f の積 $\tau_1 \cdot f_1$ は, wavelet に共通の定数



Scale representation



Melli変換により
同一の表現に帰着

基礎となる表現

Mellin Transform

$$S(p) = \int_0^{\infty} s(t) t^{p-1} dt = \int_0^{\infty} s(t) e^{(p-1)\ln t} dt$$

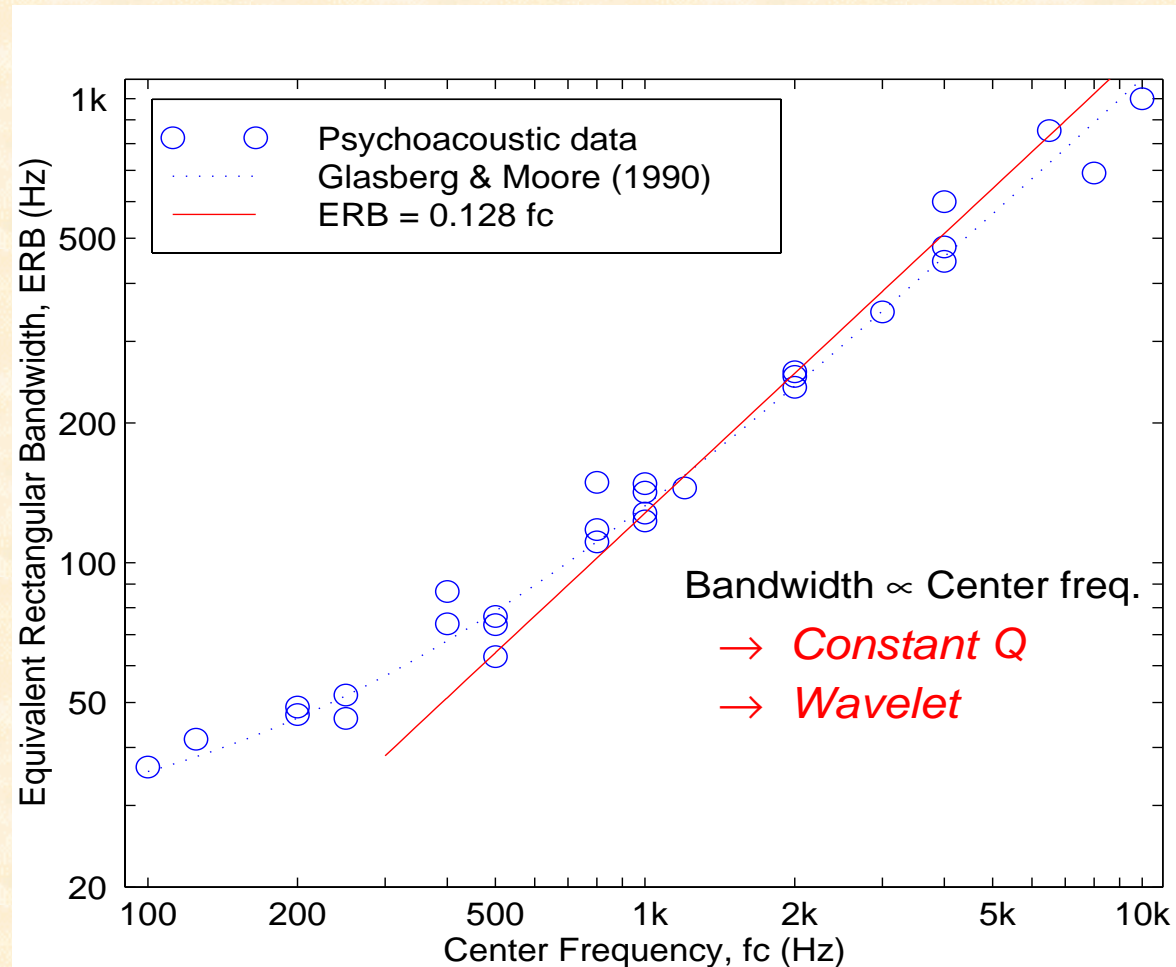
Fourier Transform

$$S(\omega) = \int_{-\infty}^{\infty} s(t) e^{-i\omega t} dt$$

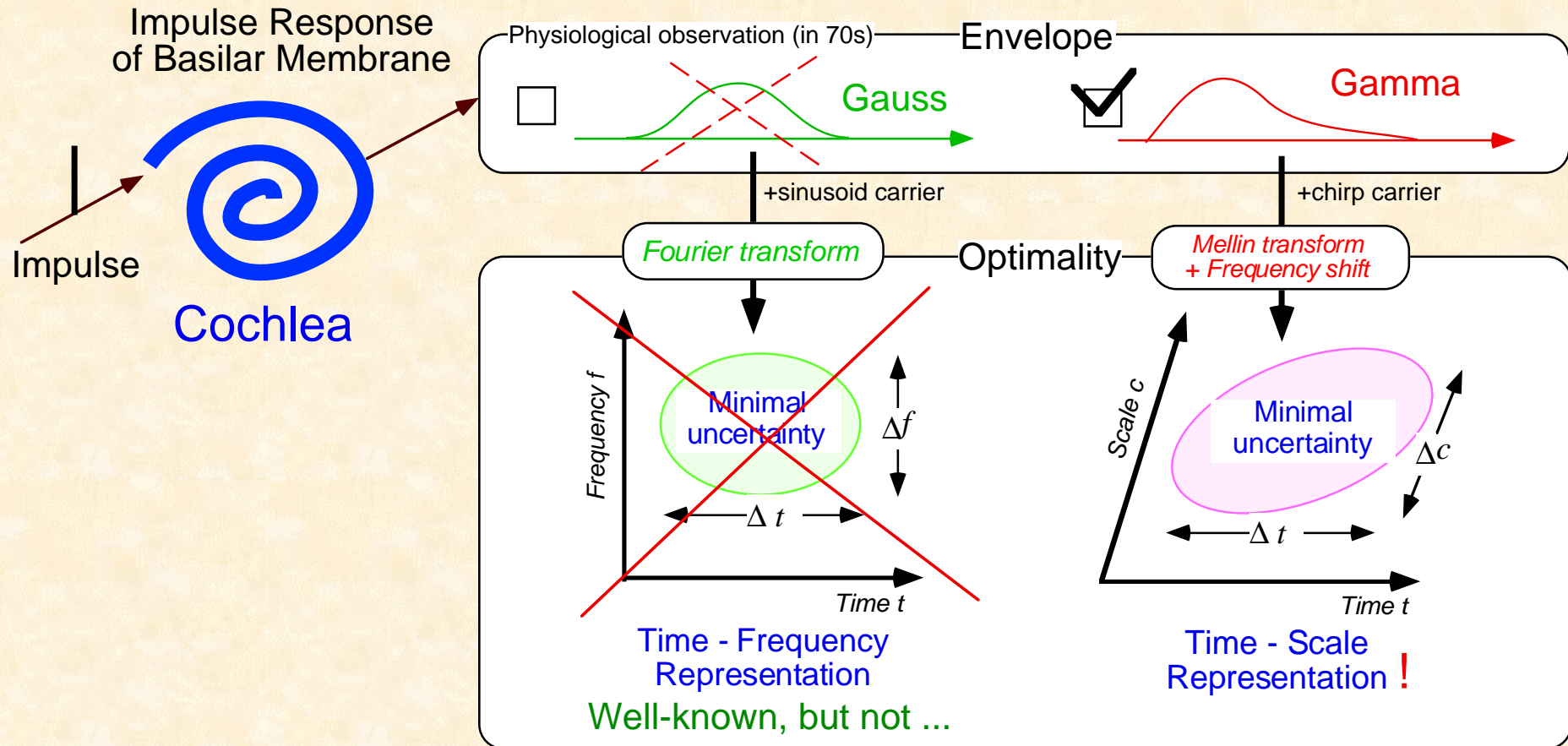
原点の指定



内耳における周波数分析



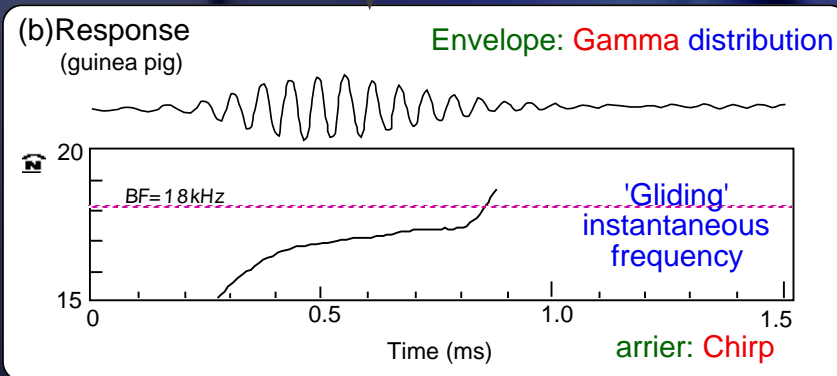
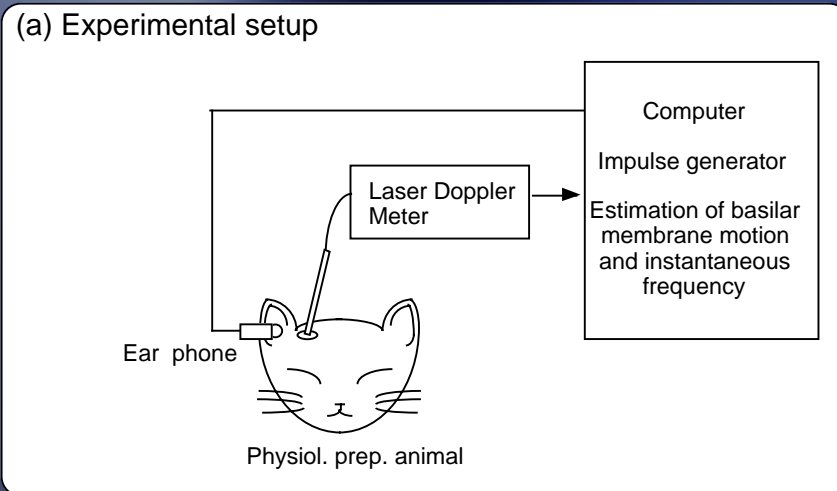
内耳における分析の最適性



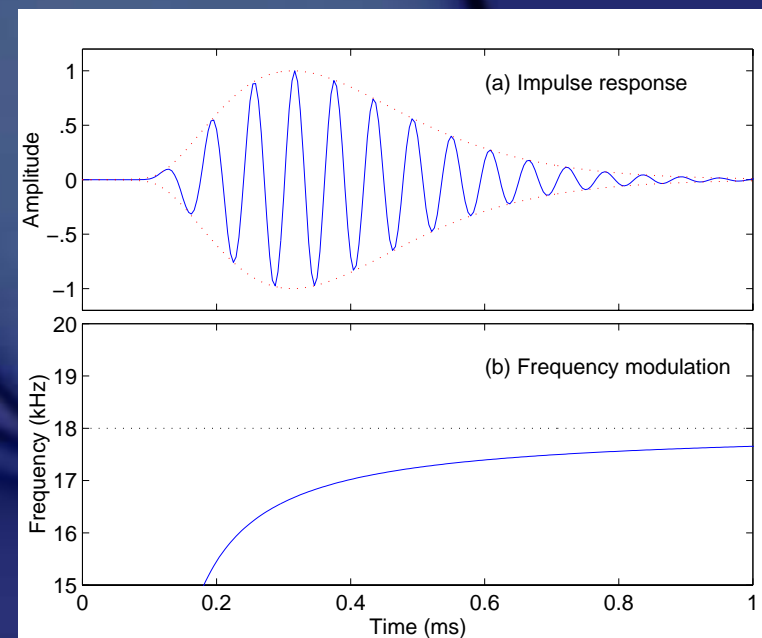
時間-スケール領域での 最小不確定性を有する関数

$$s(t) = kt^{\alpha_2} e^{-\alpha_1 t + j\langle c \rangle \ln(t / \langle t \rangle)}$$

生理学的裏付け



Gammachirp



de Boer and Nuttall (1997)

Irino (1995,1996),
Irino and Patterson (1997, 2001)

発表の概要

n デモ：何ができるようになるか？

n 聴覚に本質的な量をどう表現し求めるか？

—音色の知覚における不変性

n stabilized wavelet-Mellin transform and gammachirp

—ピッチ知覚と基本周波数

n instantaneous frequency of wavelet analysis

—音源の知覚と音響イベント

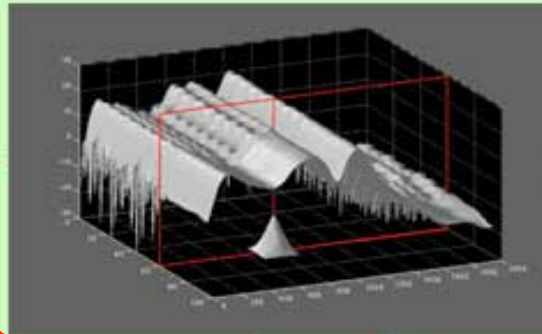
n multi-resolution zero-crossing and causality

STRAIGHT

F0-adaptive time-frequency smoothing
to eliminate periodicity interferences

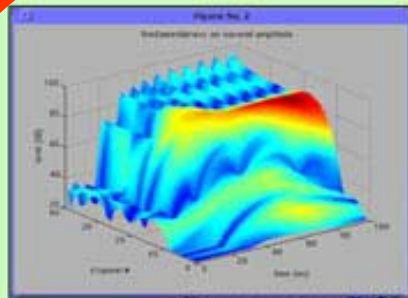
入力信号

出力音声

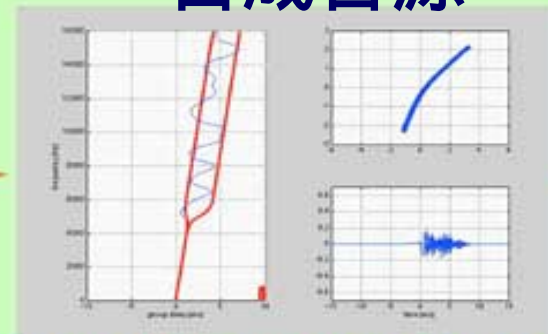


伝達特性

合成音源



音源情報：
周期性等



Instantaneous-frequency-based
F0 and source information extractor

Group delay manipulation to add
artificial naturalness

STRAIGHT is a very high-quality VOCODER.

発表の概要

n デモ：何ができるようになるか？

n 聴覚に本質的な量をどう表現し求めるか？

—音色の知覚における不変性

n stabilized wavelet-Mellin transform and gammachirp

—ピッチ知覚と基本周波数

n instantaneous frequency of wavelet analysis

—音源の知覚と音響イベント

n multi-resolution zero-crossing and causality

瞬時周波数に基づいたF0抽出

- n 変化が本質的

- n 不動点に基づく方法

 - 二段階選択

 - n 不動点とwaveletを利用したC/Nでの選択

 - 複数調波成分の利用

 - n C/Nを利用して統合，精度の改善

信号のモデル

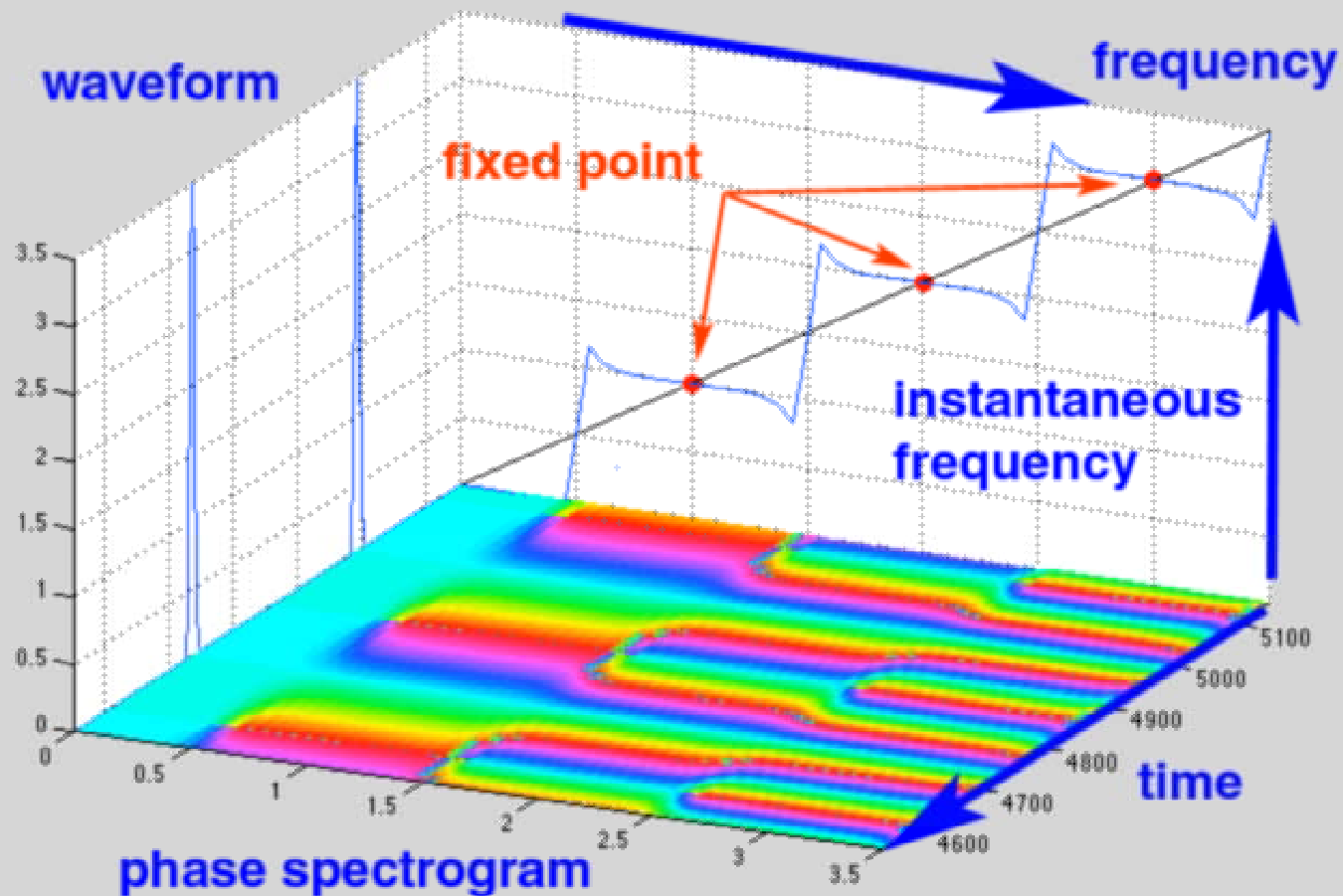
$$s(t) = \sum_{k=1}^N a_k(t) e^{j\left\{ \int_0^t \omega_k(\tau) d\tau + \phi_k(0) \right\}}$$

瞬時周波数

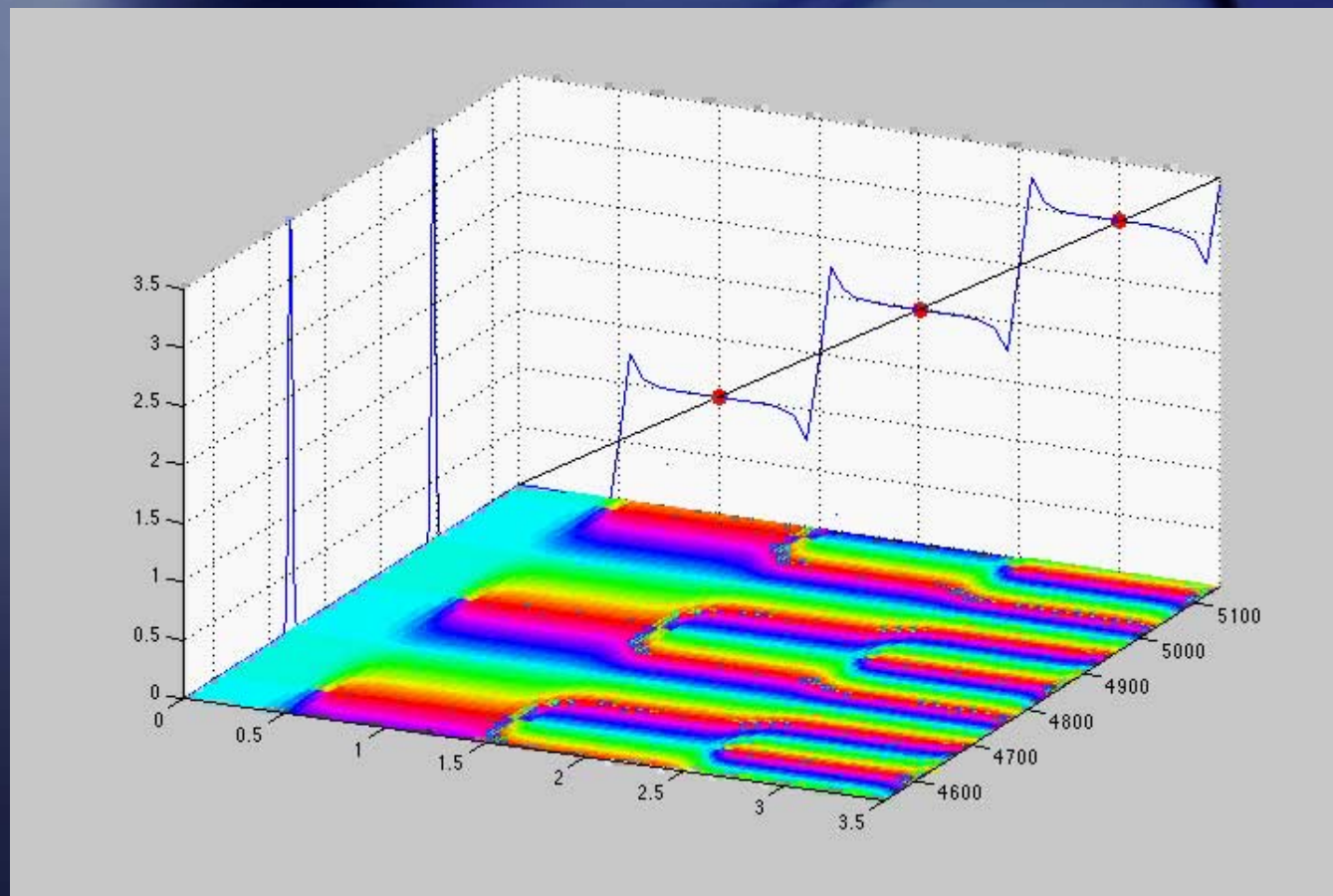
音声の場合： k に調波的構造

$k=0$: 基本波成分

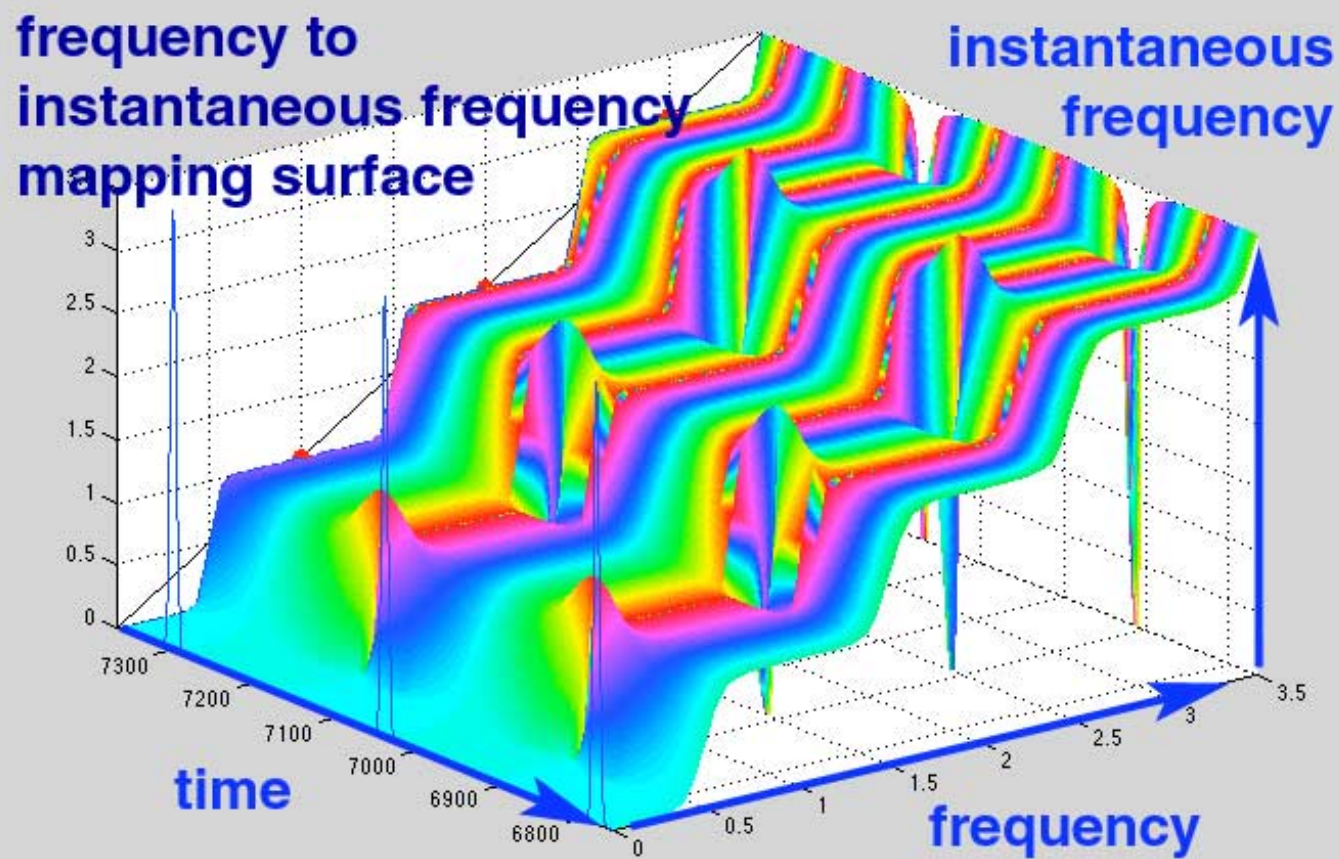
不動点による正弦波成分の選択



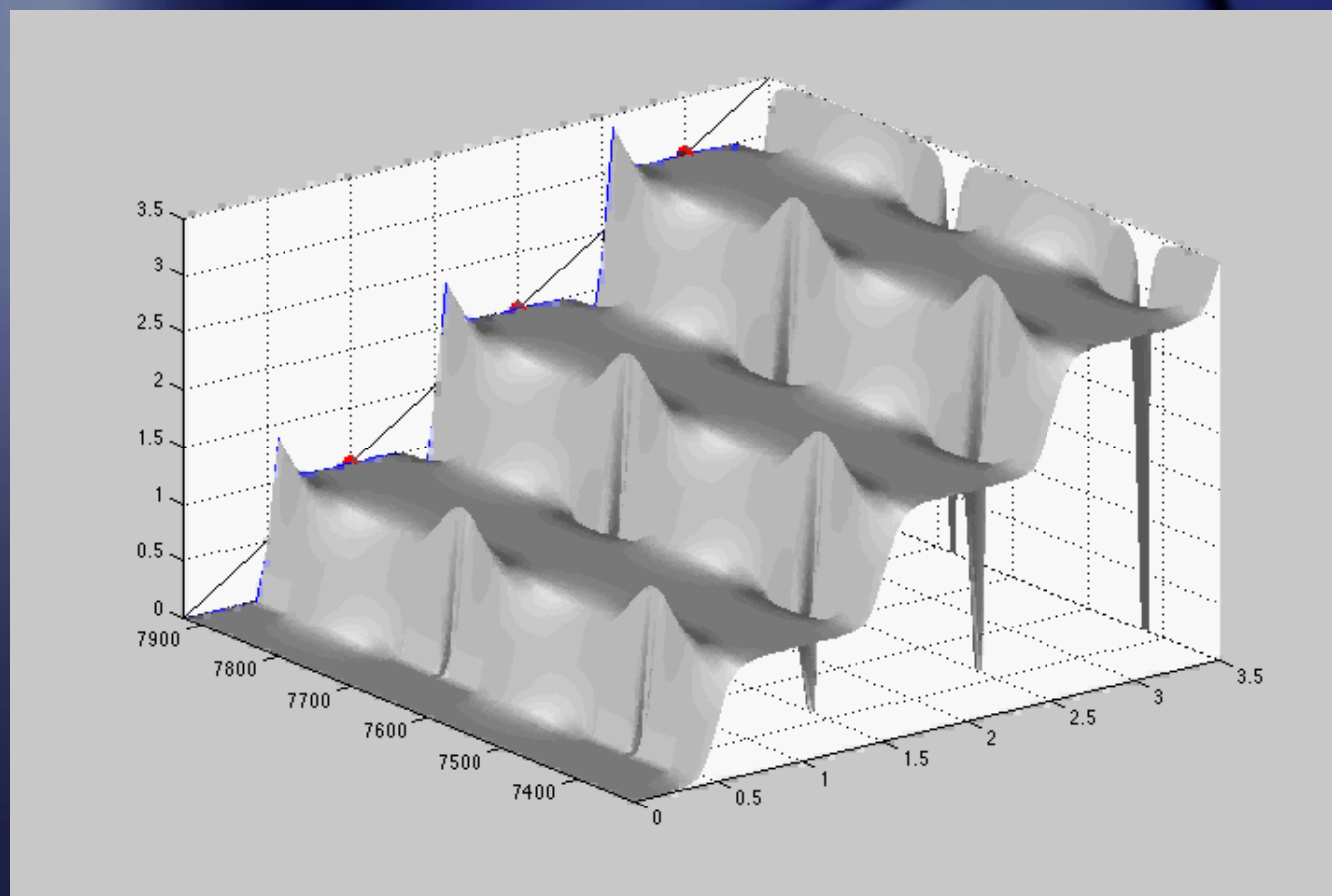
不動点による正弦波成分の選択



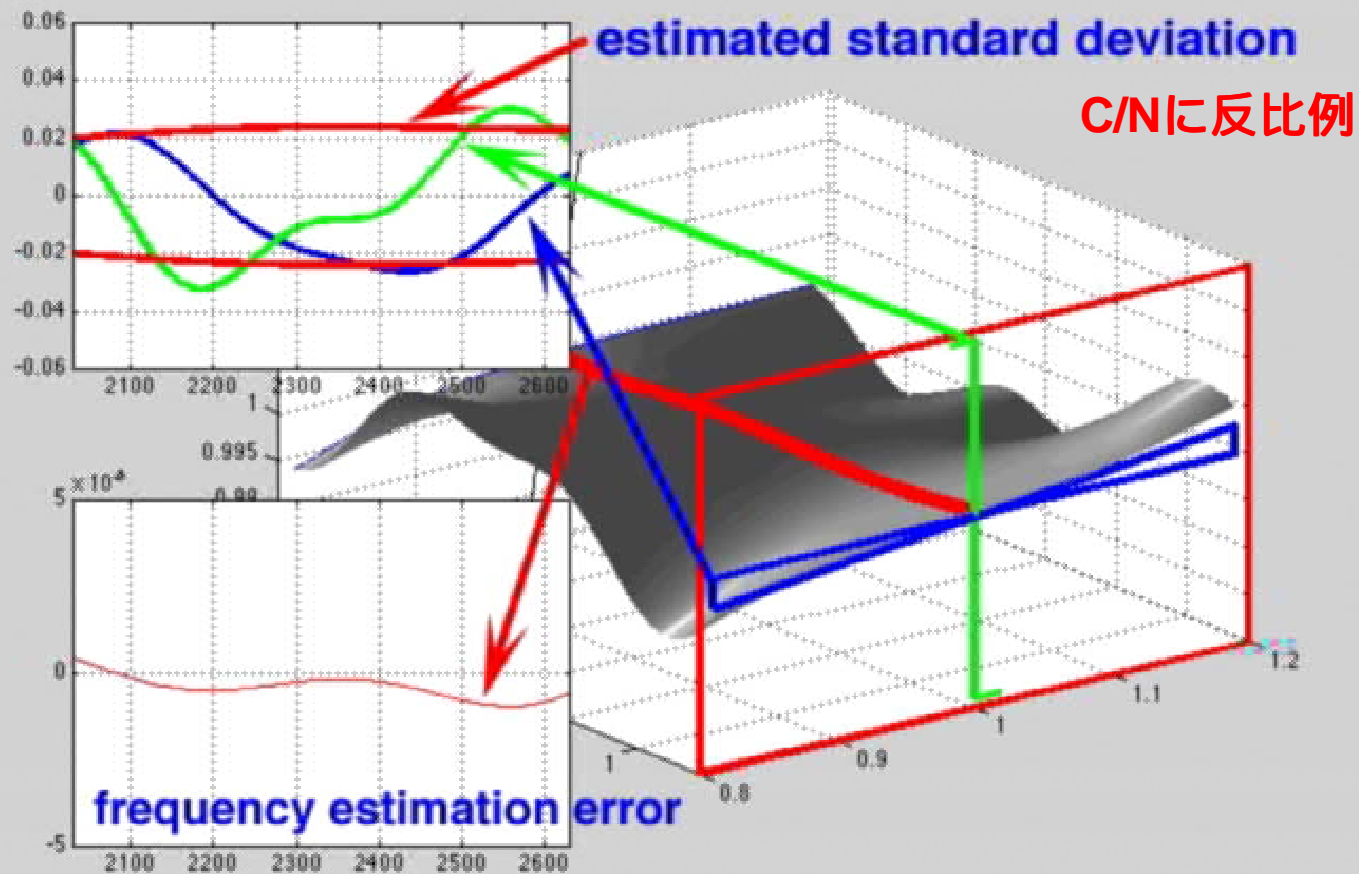
時間-周波数領域における 周波数 瞬時周波数写像



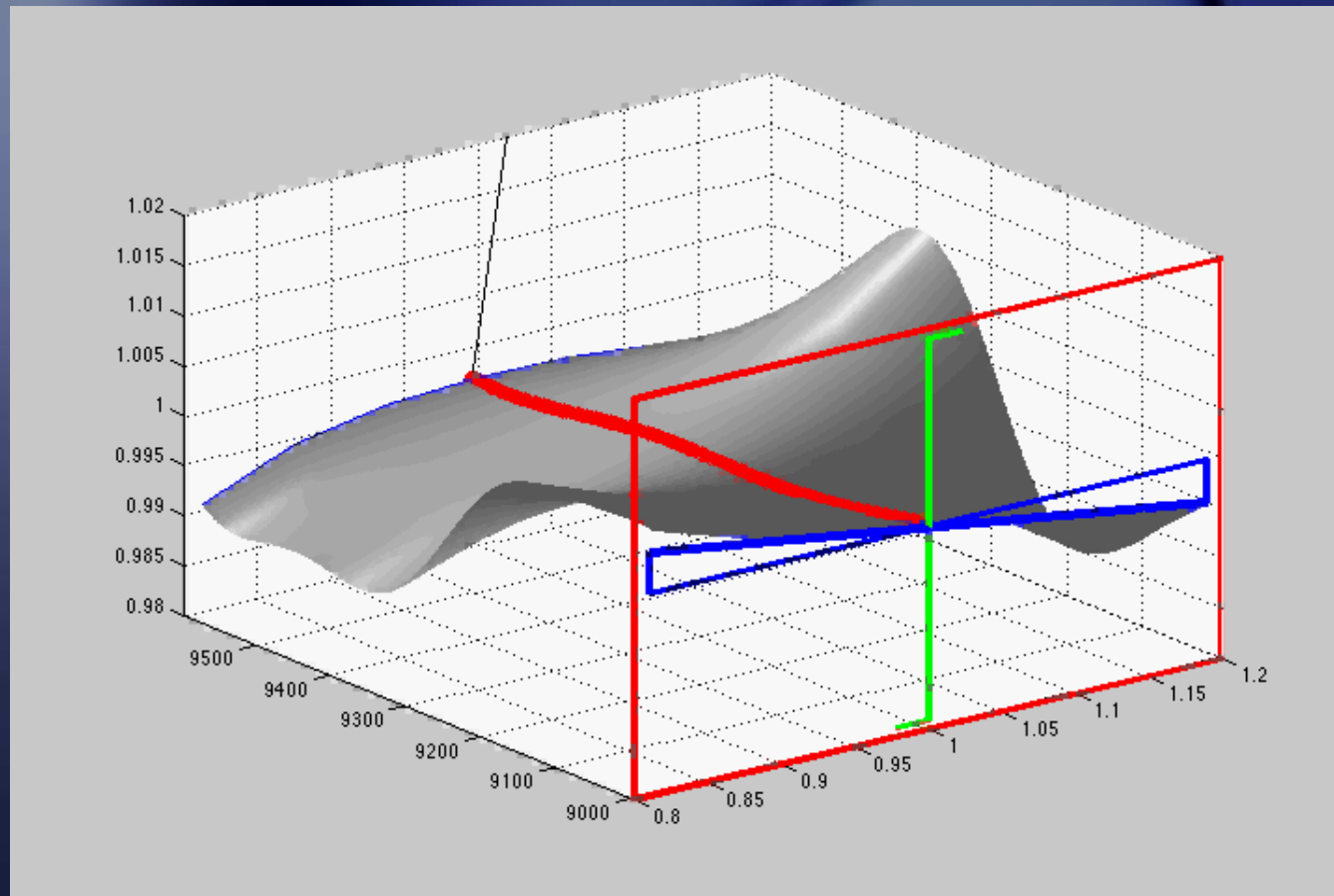
時間-周波数領域における 周波数 瞬時周波数写像



不動点近傍の写像の形状に基づく 正弦波成分のC/Nの推定



不動点近傍の写像の形状に基づく 正弦波成分のC/Nの推定



搬送周波数にチューンした窓関数

$$w_s(t, \lambda_c) = w(t, \lambda_c) * h(t, \lambda_c) ,$$

$$w(t, \lambda_c) = e^{-\frac{\lambda_c^2 t^2}{4\pi\eta^2}} e^{j\lambda_c t} ,$$

$$h(t, \lambda_c) = \max \left\{ 0, 1 - \left| \frac{\lambda_c t}{2\pi\eta} \right| \right\} ,$$

C/Nの近似推定

$$\bar{\sigma}^2(t, \lambda) = \int_{-T_w}^{T_w} |w(\tau, \lambda)| \tilde{\sigma}^2(t - \tau, \lambda) d\tau$$

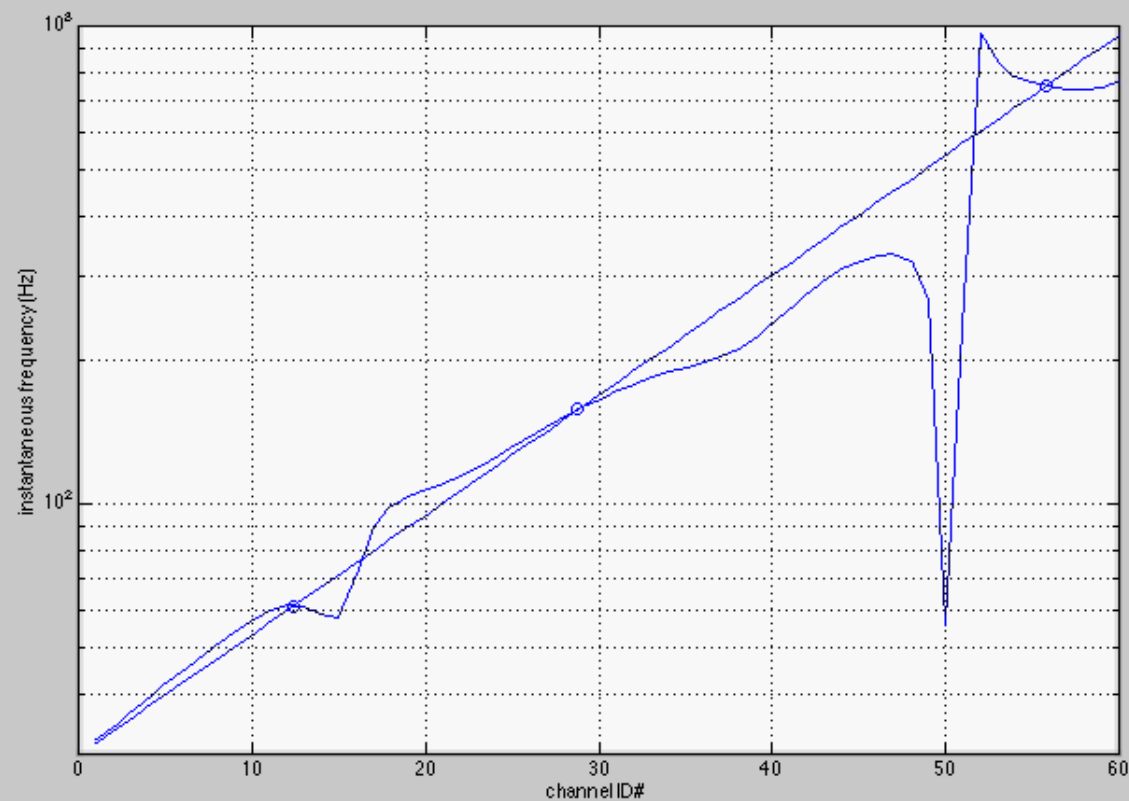
$$\tilde{\sigma}^2(t) = c_a \left(\frac{\partial \omega(t, \lambda)}{\partial \lambda} \right)^2 + c_b \left(\frac{\partial^2 \omega(t, \lambda)}{\partial t \partial \lambda} \right)^2$$

$$c_a = \frac{1}{\int_{-\infty}^{\infty} \left(\lambda_o \frac{dg(\lambda)}{d\lambda} \Big|_{\lambda=\lambda_o} \right)^2 d\lambda_o} .$$

$$c_b = \frac{1}{\int_{-\infty}^{\infty} \left(\lambda_o^2 \frac{dg(\lambda)}{d\lambda} \Big|_{\lambda=\lambda_o} \right)^2 d\lambda_o} .$$

waveletによる 基本波成分の特別扱い

フィルタ出力の瞬時周波数

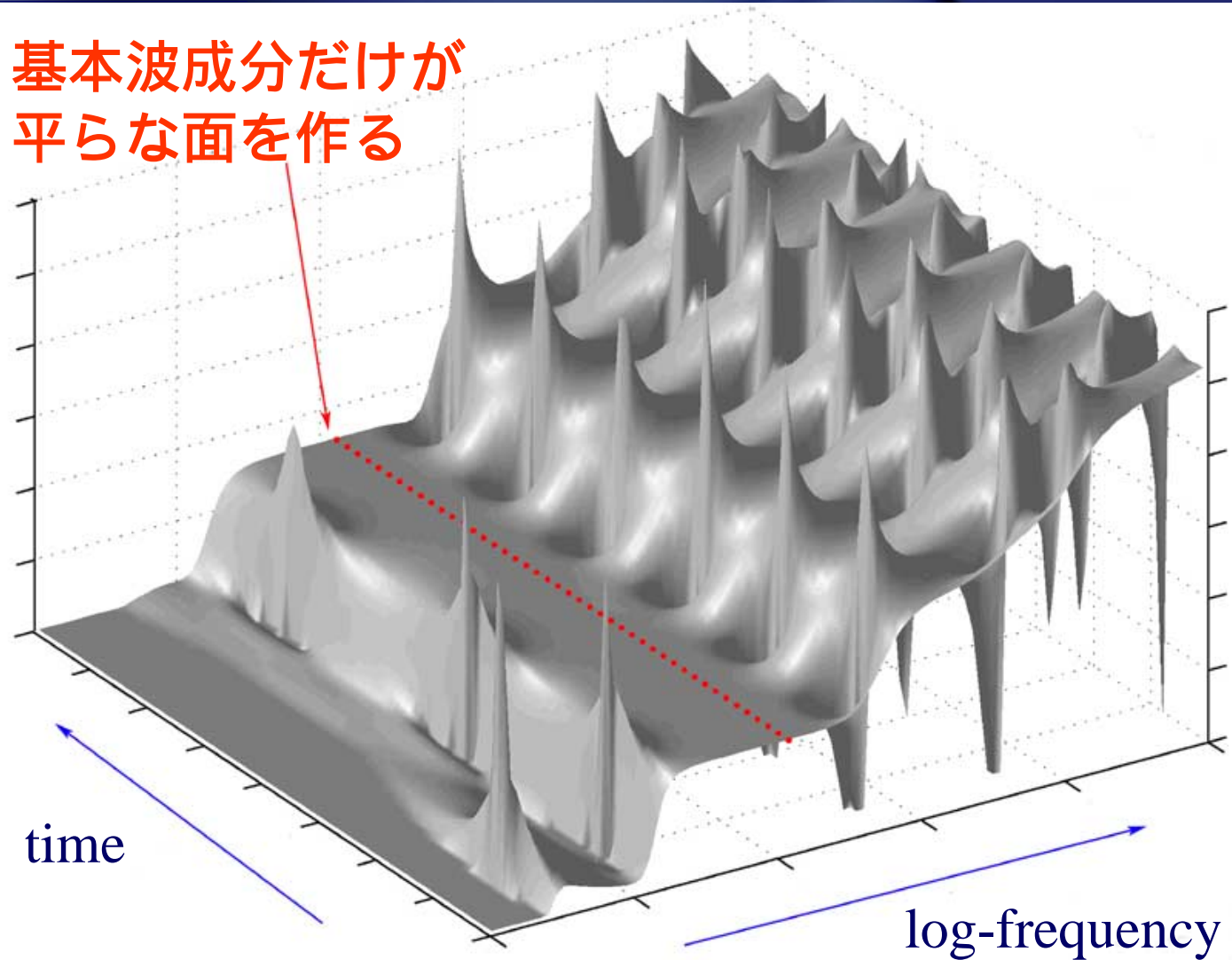


フィルタ番号：12個 / オクターブ (0番：30Hz)

log-instantaneous
frequency of
filter output

waveletによる 基本波成分の特別扱い

基本波成分だけが
平らな面を作る



STRAIGHTでの 実装

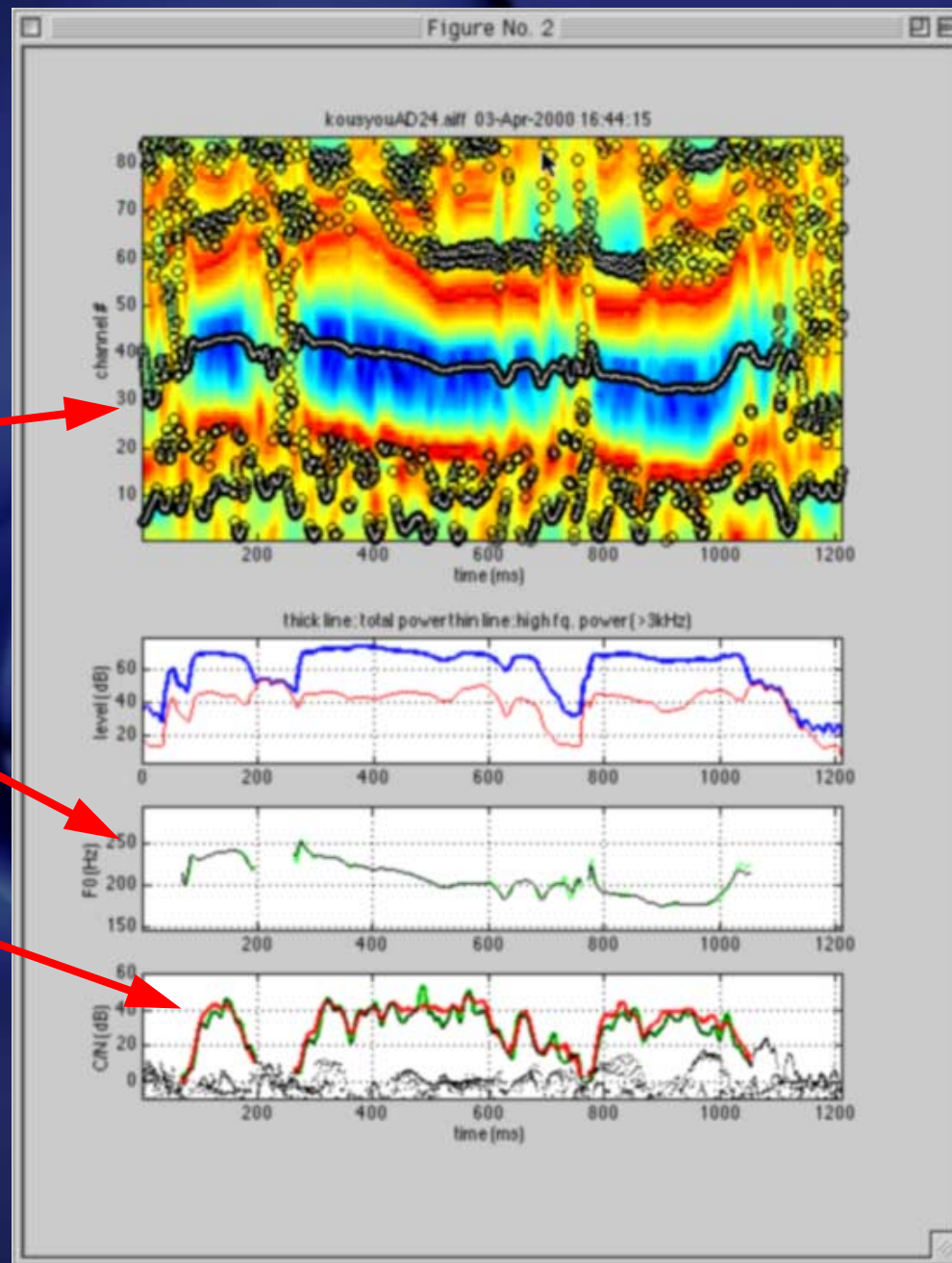
n C/N : 背景のカラー
不動点 : 印

n 抽出されたF0

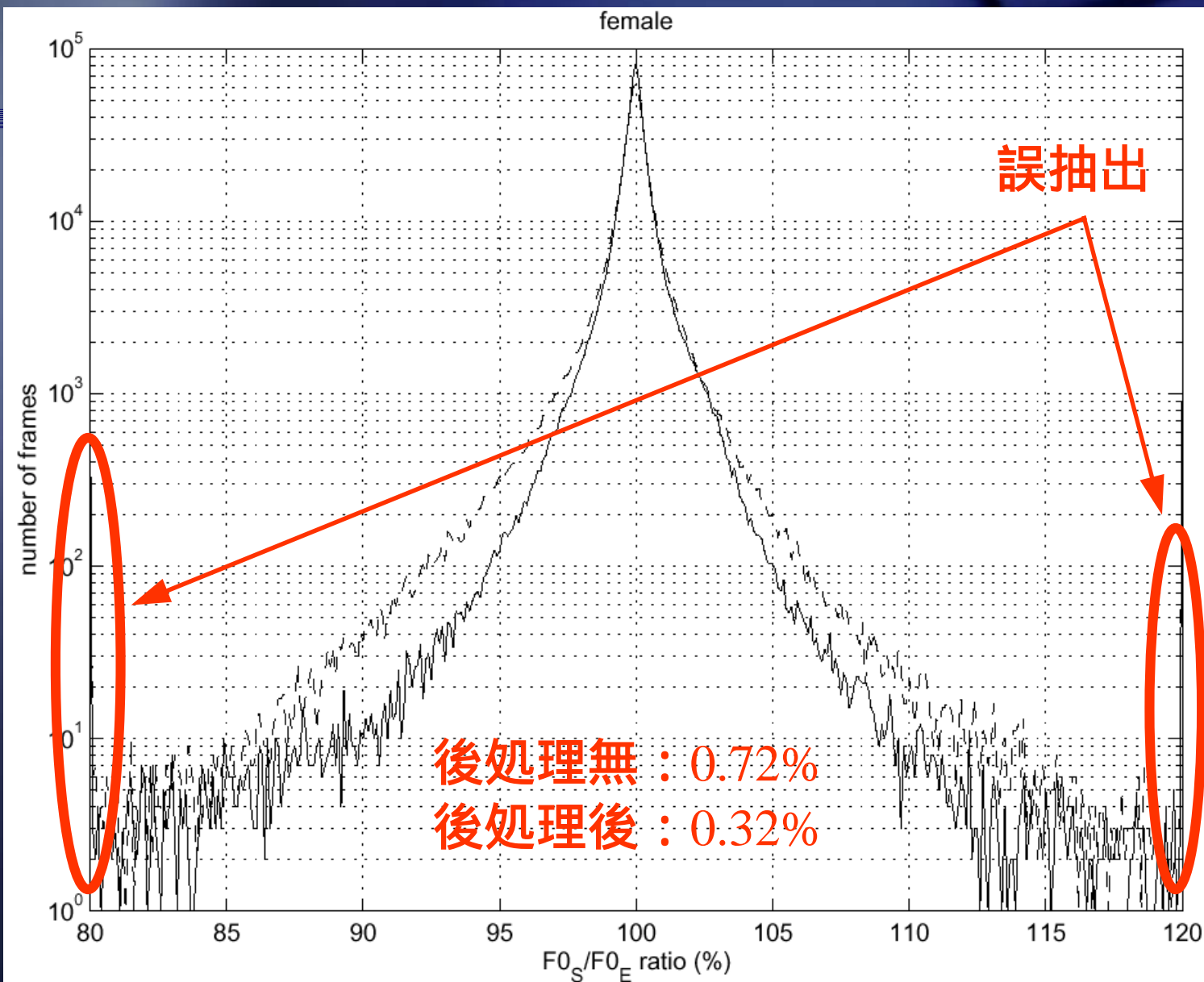
n 各不動点のC/N

—基本波成分 : 緑, 赤

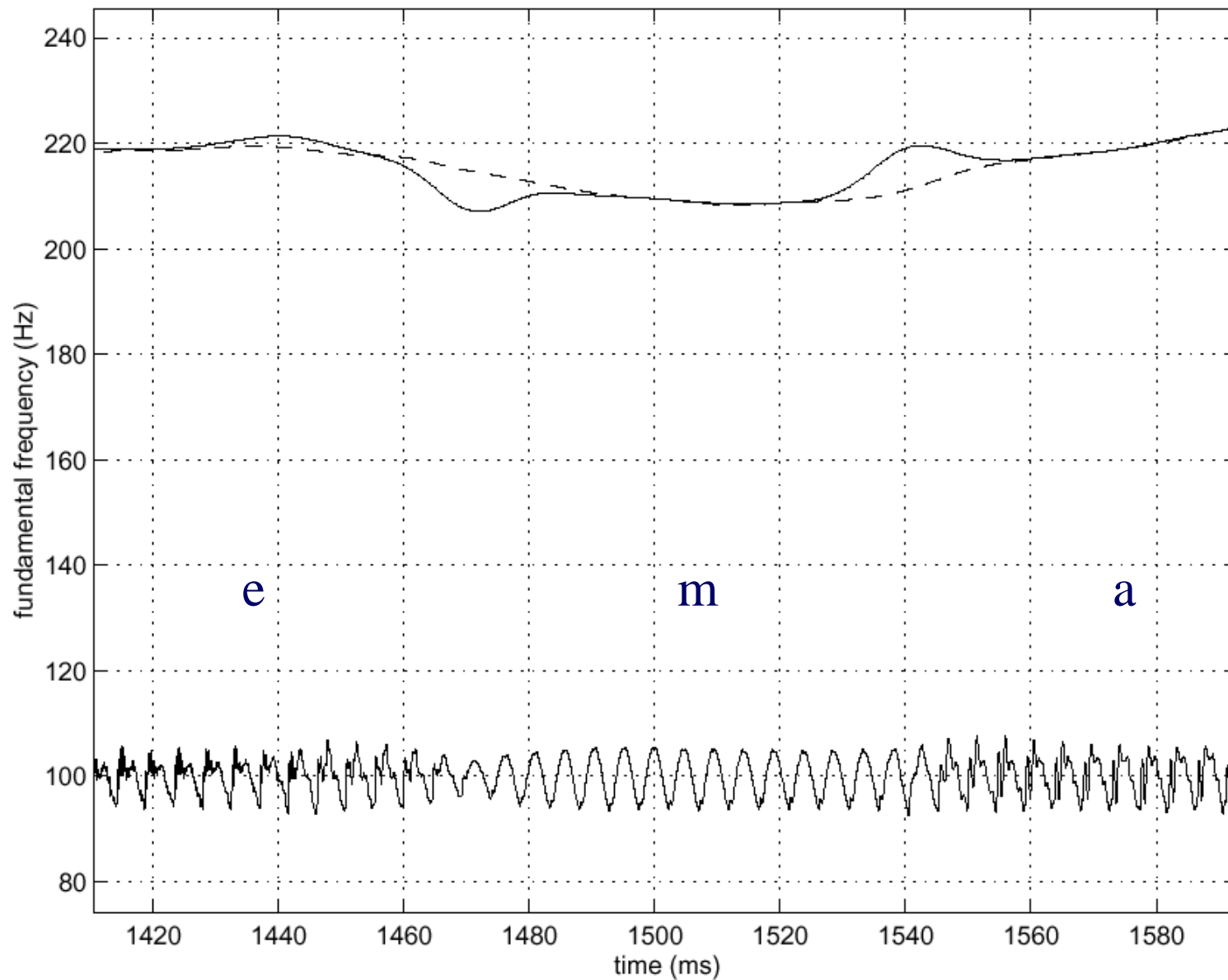
—その他の成分 : 黒



EGGを基準とした評価：女性



ドップラー効果？



発表の概要

n デモ：何ができるようになるか？

n 聴覚に本質的な量をどう表現し求めるか？

—音色の知覚における不変性

n stabilized wavelet-Mellin transform and gammachirp

—ピッチ知覚と基本周波数

n instantaneous frequency of wavelet analysis

—音源の知覚と音響イベント

n multi-resolution zero-crossing and causality

イベント：時間領域での定義

n 窓の中心時刻から窓内のエネルギーの重心位置への写像の不動点

— イベントの属性

n 位置（平均時刻）

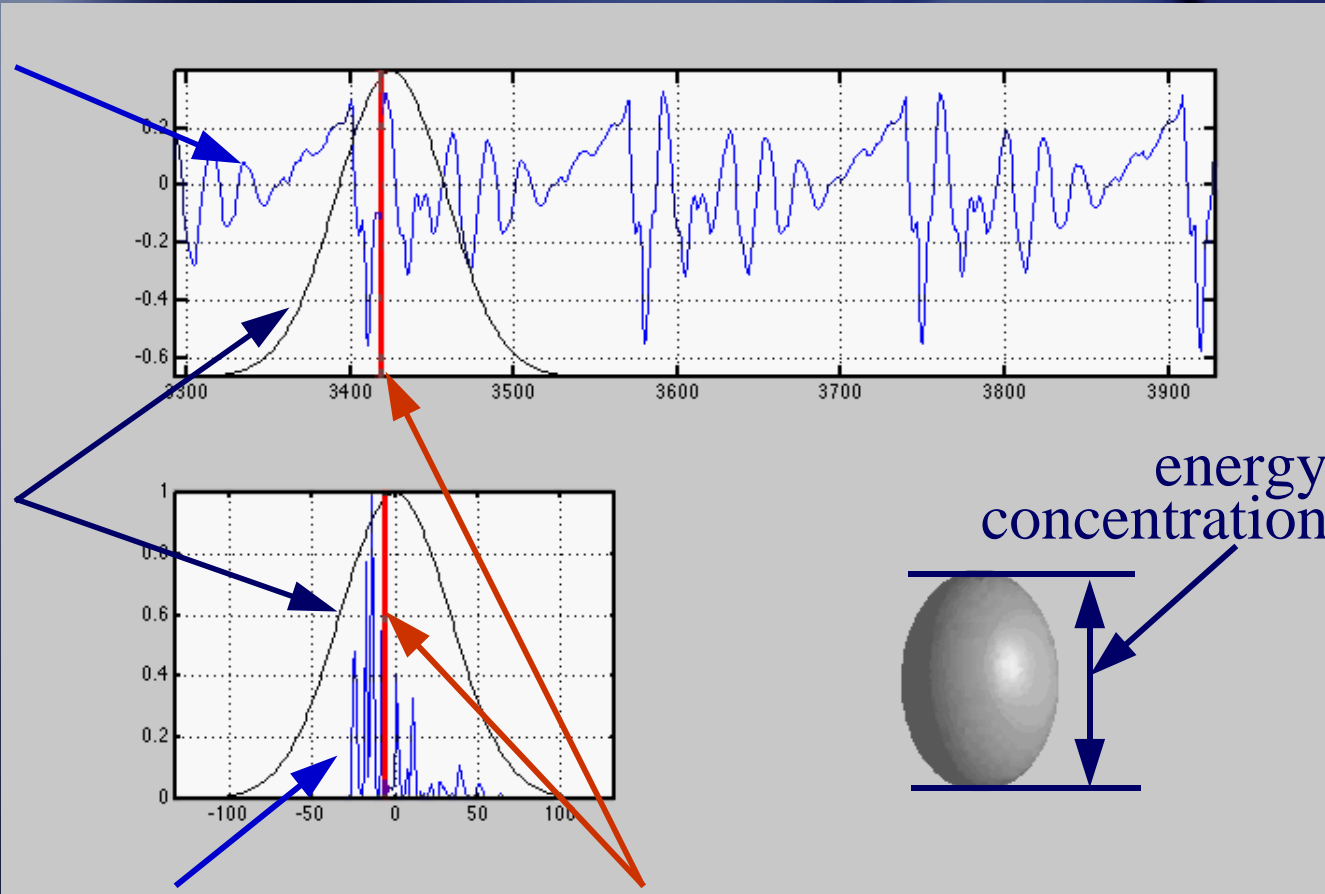
n 継続時間（標準偏差）

— パラメタ

n スケール

Acoustic event: time domain

speech waveform

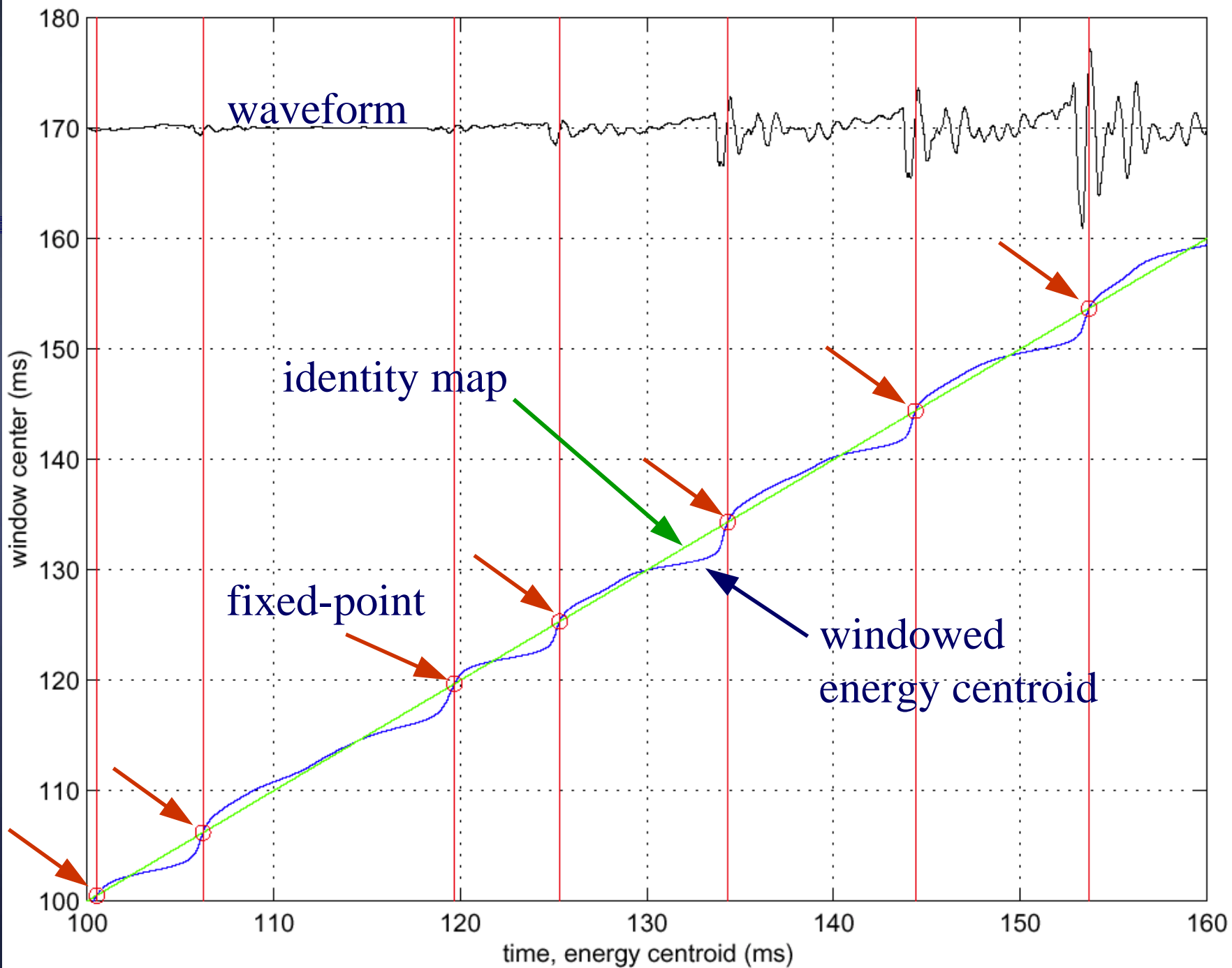


Gaussian window

energy concentration

squared whitened signal

energy centroid



Event: time-domain definition

windowed and whitened signal

$$\langle t(u) \rangle = \frac{\int t |x(t, u)|^2 dt}{\int |x(t, u)|^2 dt}$$

mean time

$$\sigma_t^2(u) = \frac{\int (t - \langle t \rangle)^2 |x(t, u)|^2 dt}{\int |x(t, u)|^2 dt}$$

duration

$$\{t_e\} = \{u | \langle t(u) \rangle = u, \frac{d\langle t(u) \rangle}{du} < 1\}$$

event locations

Event: time-domain definition

$$\langle t(u) \rangle = \frac{\int t |x(t, u)|^2 dt}{\int |x(t, u)|^2 dt}$$

mean time

$$\sigma_t^2(u) = \frac{\int (t - \langle t \rangle)^2 |x(t, u)|^2 dt}{\int |x(t, u)|^2 dt}$$

duration

$$\{t_e\} = \left\{ u \mid \langle t(u) \rangle = u, \frac{d\langle t(u) \rangle}{du} < 1 \right\}$$

event locations

Actual event location and windowed location

$$w(t) = e^{-\frac{t^2}{2\sigma_w^2}}.$$

Gaussian time window

$$|s(t)| = e^{-\frac{(t-t_e)^2}{2\sigma_s^2}},$$

Gaussian signal model

$$\langle t(u) \rangle = \frac{\sigma_s^2 u + \sigma_w^2 t_e}{\sigma_s^2 + \sigma_w^2}$$

windowed event location

window location

actual event location

Event duration estimation from mapping gradient at fixed points

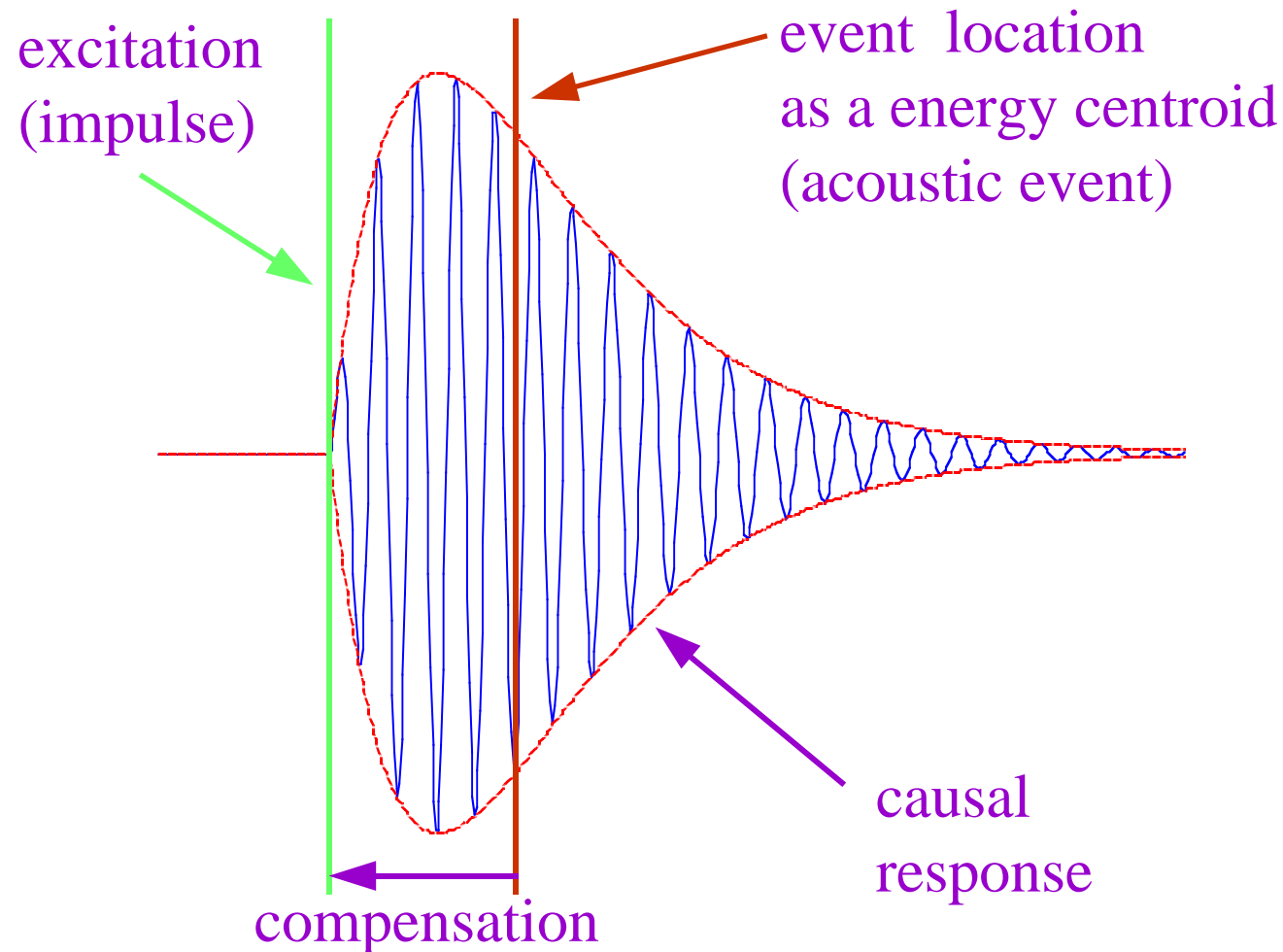
$$\sigma_s(t_e) = \sigma_w \sqrt{\frac{g(t_e)}{1 - \underline{g(t_e)}}}$$

event duration

scale

gradient of mapping
at fixed points

From acoustic event to its excitation: inverse problem



From acoustic event to its excitation: inverse problem

- n Frequency domain representations of event attributes defined in the time-domain

$$\langle t(u) \rangle = - \int \psi'(\omega, u) |S(\omega, u)|^2 d\omega$$

$$\sigma_t^2(u) = \int \left(\frac{B'(\omega, u)}{B(\omega, u)} \right)^2 B^2(\omega, u) d\omega + \int (\psi'(\omega, u) + \langle t(u) \rangle)^2 B^2(\omega, u) d\omega$$

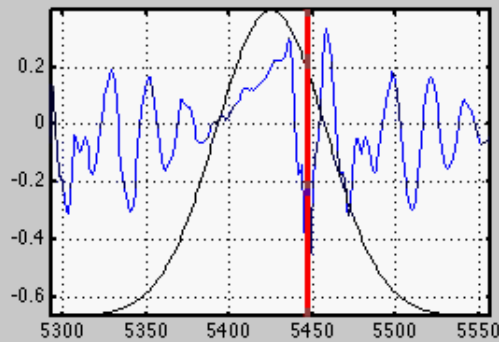
$$S(\omega, u) = \frac{1}{\sqrt{2\pi}} \int x(t, u) e^{-j\omega t} dt = |S(\omega, u)| e^{j\psi(\omega, u)} = B(\omega, u) e^{j\psi(\omega, u)}$$

負号をつけたものが群遅延

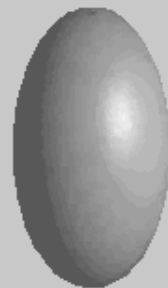
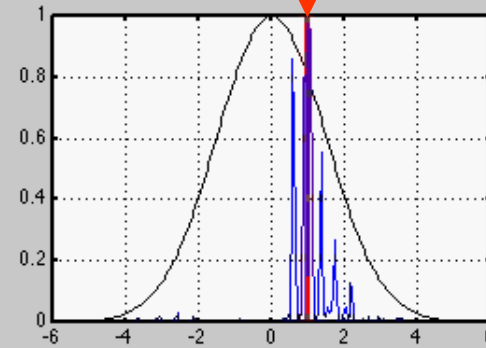
- n Assumption: causality

Equivalence of temporal and frequency representations

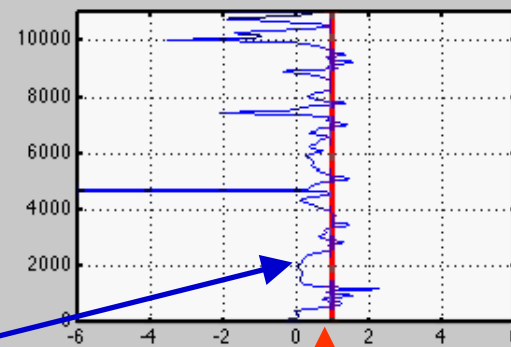
時間領域で求めた重心



波形



群遅延



周波数領域で求めた重心

因果システムの群遅延

$$\tau_{\phi}(\omega, u) = -\frac{d}{d\omega} \left(\text{imag} \left[\frac{1}{\sqrt{2\pi}} \int C(q, u) e^{j\omega q} dq \right] \right) \quad (14)$$

$$C(q, u) = \begin{cases} 2c(q, u) & q > 0 \\ c(q, u) & q = 0 \\ 0 & \text{その他の場合} \end{cases}$$

$$c(q, u) = \frac{1}{\sqrt{2\pi}} \int \log B(\omega, u) e^{-j\omega q} d\omega \quad (15)$$

Excitation estimation by minimum phase compensation

$$\langle \tilde{t}(u) \rangle = - \int (\psi'(\omega, u) + \tau_\phi(\omega, u)) |S(\omega, u)|^2 d\omega$$

compensated location

$$\tilde{\sigma}_P^2(u) = \int (\psi'(\omega, u) + \langle \tilde{t}(u) \rangle + \tau_\phi(\omega, u))^2 |S(\omega, u)|^2 d\omega$$

compensated duration

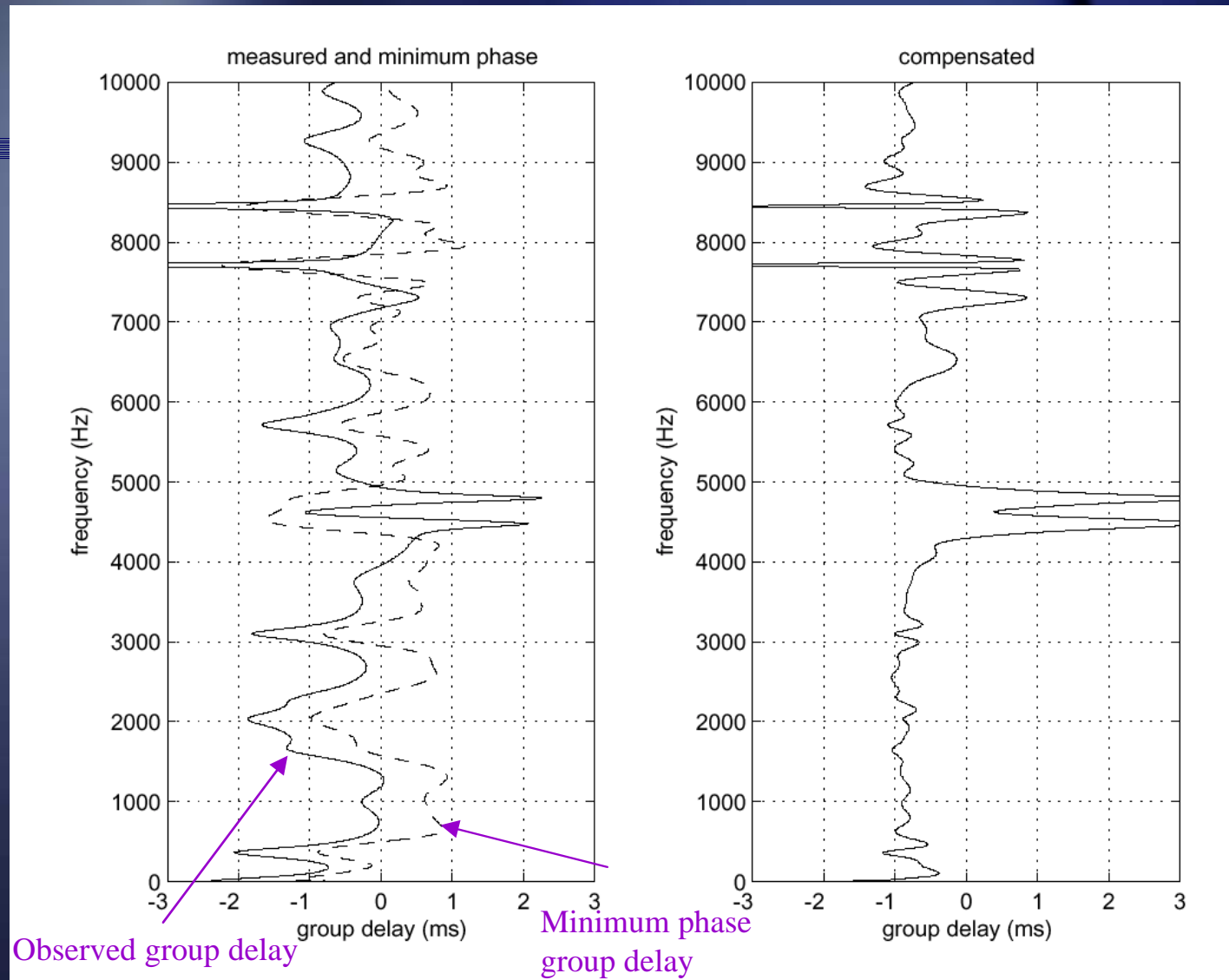
$$-\psi'(\omega, u)$$

observed group delay

$$\tau_\phi(\omega, u)$$

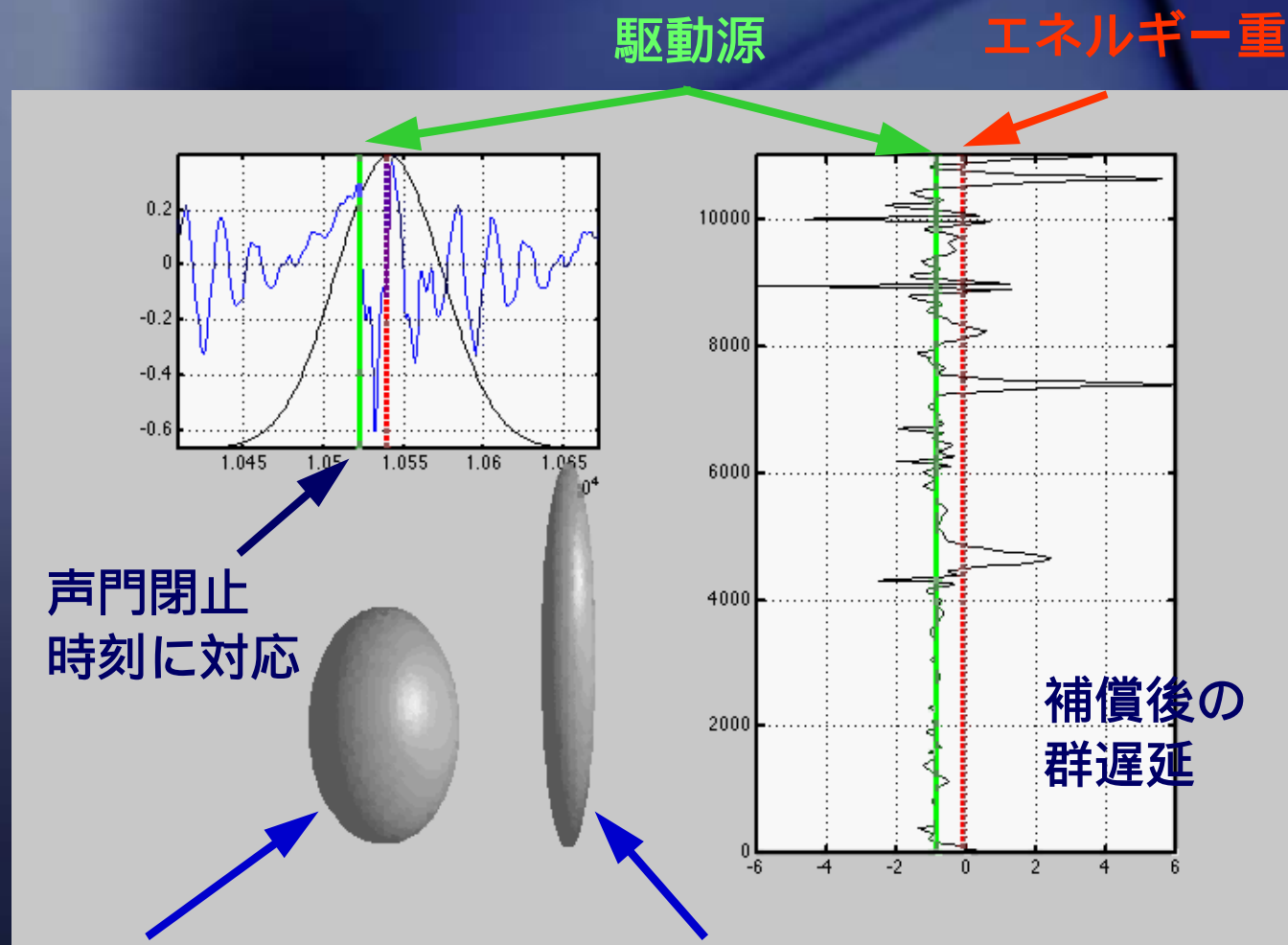
minimum phase group delay

Observed group delay, minimum phase group delay and compensated group delay



Compensated group delay

イベント抽出と駆動源の推定



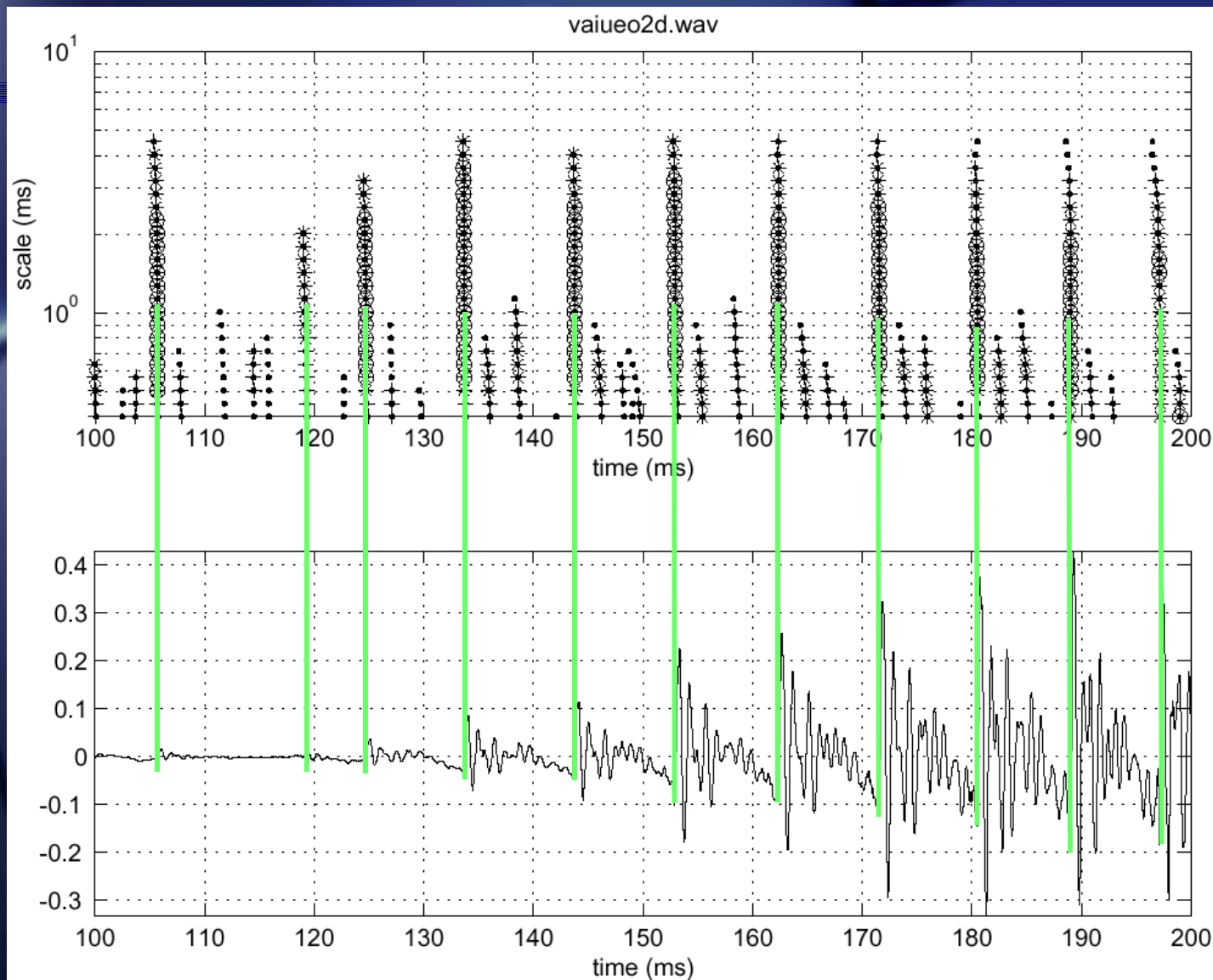
イベントのエネルギー集中度

駆動源のエネルギー集中度

イベントの多重解像度分析



推定された
駆動源位置



音声波形

まとめ：聴覚とwavelet

n waveletの必然性

—生態学 相似変換での不変性の表現

n 安定化 wavelet-Mellin 変換

—時間-スケール領域での最小不確定性

n 従来技術を凌駕するアルゴリズム

—基本周波数の抽出

—イベントと駆動源情報の抽出

n 聴覚システム理解への新しい視点の提案

—聴覚神経の同期発火のチャンネル間同期の役割