月4 2013 冬学期 [4830-1032]

第5回

音声音響信号処理

（時間周波数解析）

亀岡弘和

東京大学大学院情報理工学系研究科
日本電信電話株式会社
NTTコミュニケーション科学基礎研究所

講義内容(キーワード)

- 信号処理、符号化、標準化の実用システム例の紹介
- 情報通信の基本(誤り検出、訂正符号、変調、IP)
- 符号化技術の基本(量子化、予測、変換、圧縮)
- 音声分析・合成・認識・強調、音楽信号処理
- 統計的信号処理の基礎(スペクトル、ガウス過程、最尤推定)
- ガウス性確率変数の基本性質
- 時間周波数分析(短時間フーリエ変換、ウェーブレット変換)
- ウィナーフィルタとカルマンフィルタ
- 音声生成過程のモデル(ソースフィルタ理論と藤崎モデル)
- 自己回帰モデルと線形予測分析
- 独立成分分析によるブラインド音源分離
- 非負値行列因子分解によるスペクトログラムの分解表現
- スペクトル間擬距離
- 最適化アルゴリズム(EMアルゴリズム、補助関数法)

講義スケジュール

10/ 7	守谷先生担当
10/15 (火)	守谷先生担当
10/21	守谷先生担当
10/28	休講
11/ 5 (火)	線形予測分析と自己回帰モデル
11/11	時間周波数解析
11/18	
11/25	
12/ 2	
12/ 9	
12/16	
12/23	
1/13	
1/20	
1/27	

成績評価

- レポート課題
  - 本講義に関連する論文を1つ選び、発表資料形式(パワーポイント等)にまとめて学期末に提出してください。提出先は最終講義にてお知らせします。
  - 「どの程度本質を理解しているか」「要点が分かりやすく記述されているか」「なぜその論文を重要と考えたか」を評価の規準にして採点します。
  - 毎回の講義後にその回の講義に関連する論文を1つ挙げる予定です。それらの中から選んでも良いですし、自分で自由に探してきてもOKです。
- 講義の感想
  - レポートとともに講義に対する感想文も一緒に提出して下さい。

※講義資料は講義用ホームページにアップしていく予定。

講義URL

- http://hil.t.u-tokyo.ac.jp/~kameoka/SAP/

本日の話題

- 時間周波数解析
  - 信号を構成する周波数成分がどのように時間変化していくかを捉えるための処理
  - 近年の音声音響信号処理の研究では不可欠な要素技術（音声認識・音源分離・雑音除去・自動採譜などの前段処理としてほぼ例外なく用いられる）
  - 人間の聴覚システムでも時間周波数解析が行われていると考えられている

- 代表的な解析手法
  - 短時間Fourier変換 (Short Time Fourier Transform)
  - ウェーブレット変換（定Qフィルタバンク）

時間周波数解析

- 動機について
- 短時間Fourier変換 (Short Time Fourier Transform)
  - 定義
  - スペクトログラムとは
  - フィルタバンクとしての見方
- 聴覚フィルタバンク
  - 聴覚システムにおける時間周波数解析
  - 蝸牛モデル
- ウェーブレット変換（定Qフィルタバンク）
  - 定義
  - フィルタバンクとしての見方
- スペクトル漏れ
- スペクトログラムから信号への逆変換

時間周波数解析

- 動機について
- 短時間Fourier変換 (Short Time Fourier Transform)
  - 定義
  - スペクトログラムとは
  - フィルタバンクとしての見方
- 聴覚フィルタバンク
  - 聴覚システムにおける時間周波数解析
  - 蝸牛モデル
- ウェーブレット変換（定Qフィルタバンク）
  - 定義
  - フィルタバンクとしての見方
- スペクトル漏れ
- スペクトログラムから信号への逆変換

時間周波数解析の動機

- Fourier変換：

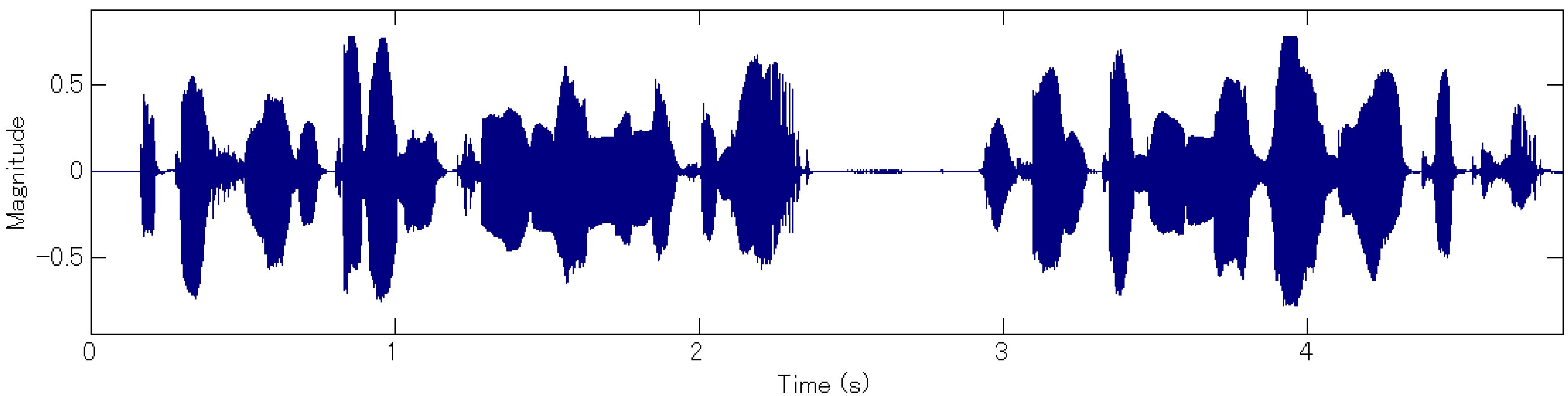
周波数ωの複素正弦波との内積

$$X(\omega) = \langle x(t), e^{j\omega t} \rangle_{t \in \mathbb{R}} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x(t) e^{-j\omega t} \mathrm{d}t$$

  - $|X(\omega)|$：信号 $x(t)$ に周波数 ω の成分がどれだけ含まれるか

→信号がどういう周期の成分から成っているかを見るのに便利

- 音声などの実環境音響信号は一般に非定常



  - 周波数成分は時々刻々と変化
  - 各時刻周辺での周波数成分を調べたい

時間周波数解析

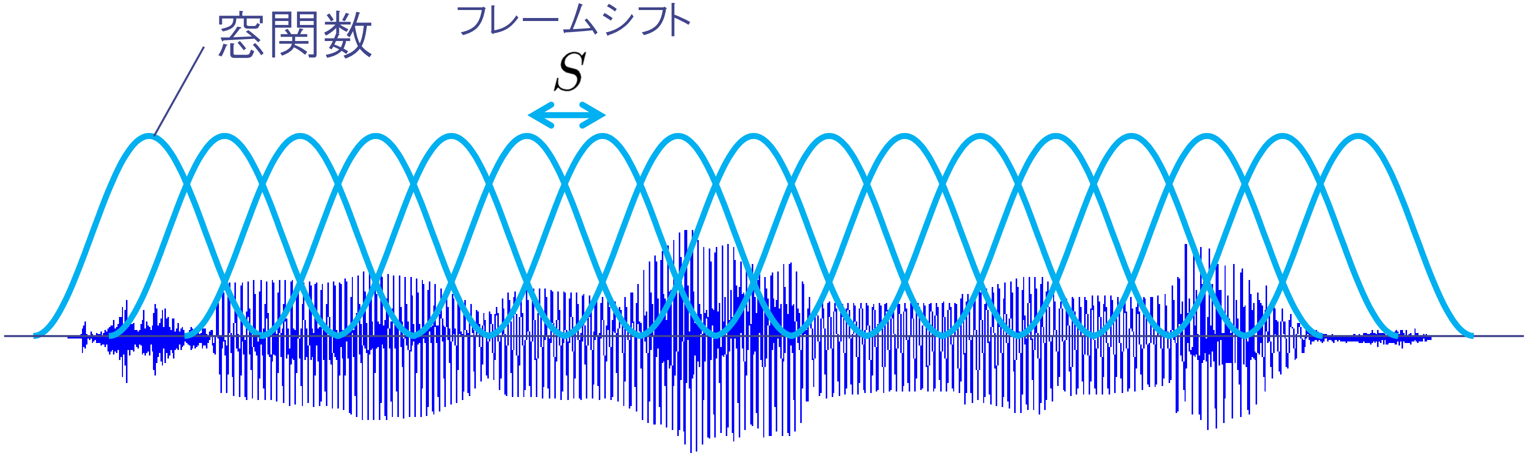
- 動機について
- 短時間Fourier変換 (Short Time Fourier Transform)
  - 定義
  - スペクトログラムとは
  - フィルタバンクとしての見方
- 聴覚フィルタバンク
  - 聴覚システムにおける時間周波数解析
  - 蝸牛モデル
- ウェーブレット変換（定Qフィルタバンク）
  - 定義
  - フィルタバンクとしての見方
- スペクトル漏れ
- スペクトログラムから信号への逆変換

時間周波数解析

- 動機について
- 短時間Fourier変換 (Short Time Fourier Transform)
  - 定義
  - スペクトログラムとは
  - フィルタバンクとしての見方
- 聴覚フィルタバンク
  - 聴覚システムにおける時間周波数解析
  - 蝸牛モデル
- ウェーブレット変換（定Qフィルタバンク）
  - 定義
  - フィルタバンクとしての見方
- スペクトル漏れ
- スペクトログラムから信号への逆変換

短時間Fourier変換（Short Time Fourier Transform）

- 文字通り，信号を短時間ごとに窓掛けして，Fourier変換する処理



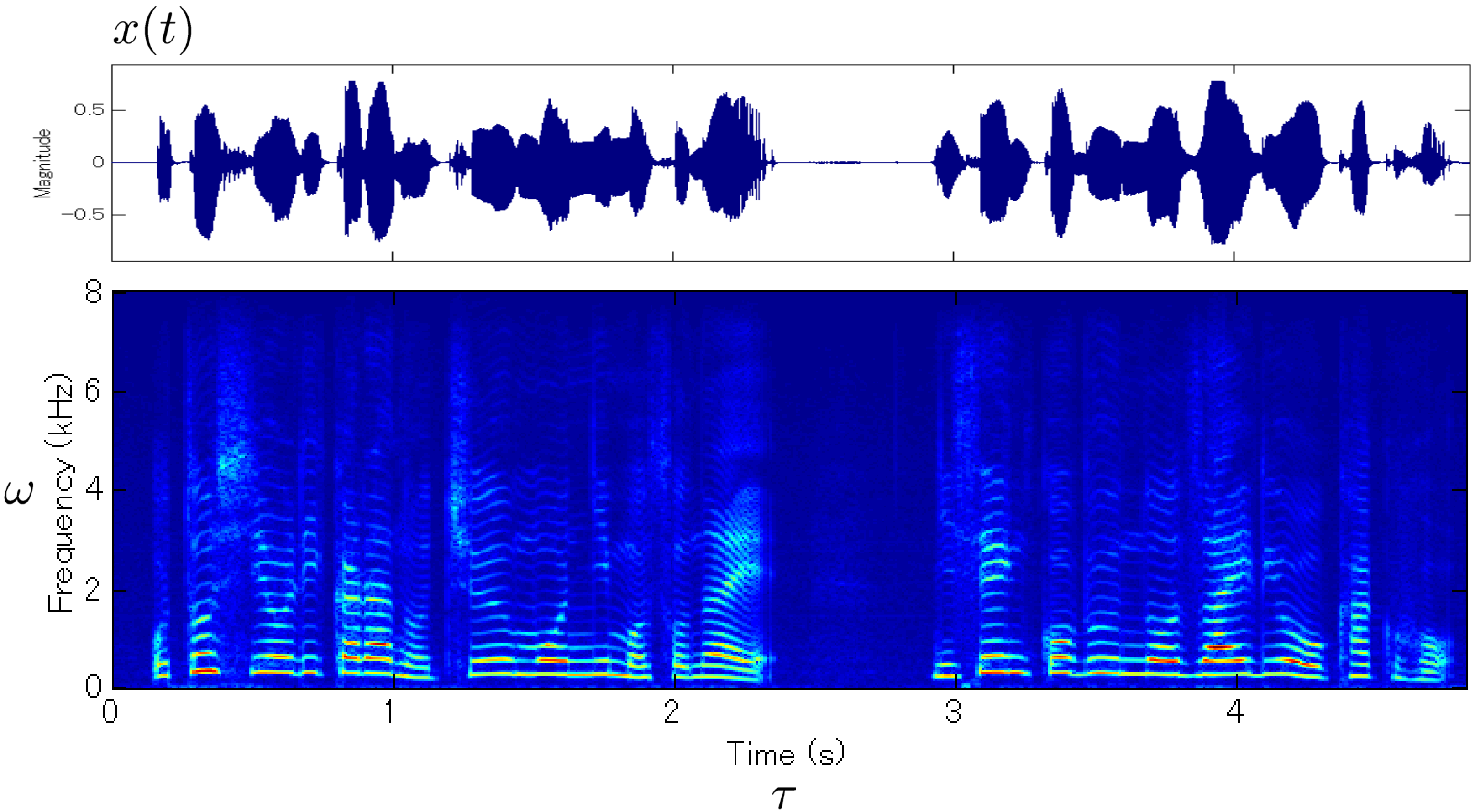
- 式で書くと・・・

m番目の窓で
切り出された波形

$$X_{\mathrm{STFT}}(\omega, mS) = \frac{1}{\sqrt{2\pi}} \int_{\infty}^{\infty} \underline{w(t)x(t+mS)} e^{-j\omega t} \mathrm{d}t$$

スペクトログラム（信号の時間周波数表現）

- $|X_{\mathrm{STFT}}(\omega, \tau)|$ をカラーマップ表示してみる



フィルタバンクとしての見方 (1/2)

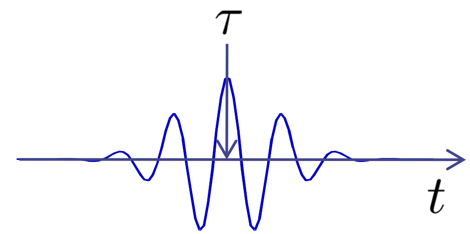
$$X_{\mathrm{STFT}}(\omega,\tau) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} w(t)x(t+\tau)e^{-j\omega t}\mathrm{d}t$$

時刻 τ を中心とした窓で切り出された波形

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \underline{w(t-\tau)x(t)}e^{-j\omega(t-\tau)}\mathrm{d}t$$

時刻 τ に局在する周波数 ω の局在波

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x(t)\underbrace{w(t-\tau)e^{-j\omega(t-\tau)}}_{\psi_{\omega,\tau}(t)}\mathrm{d}t$$



$$= \langle x(t), \psi_{\omega,\tau}(t)\rangle_{t\in\mathbb{R}}$$

$$= \langle \underline{X(y)}, \underline{\Psi_{\omega,\tau}(y)}\rangle_{y\in\mathbb{R}}$$

x と $\psi_{\omega,\tau}$ のFourier変換

一般化Parsevalの定理：
時間領域の内積は
周波数領域の内積と等しい

$$= \int_{-\infty}^{\infty} X(y)\Psi^*_{\omega,\tau}(y)\mathrm{d}y$$

$$= \int_{-\infty}^{\infty} X(y)W(y-\omega)e^{jy\tau}\mathrm{d}y$$

$$\because\ \Psi_{\omega,\tau}(y) = \Psi_{\omega,0}(y)e^{-jy\tau}$$

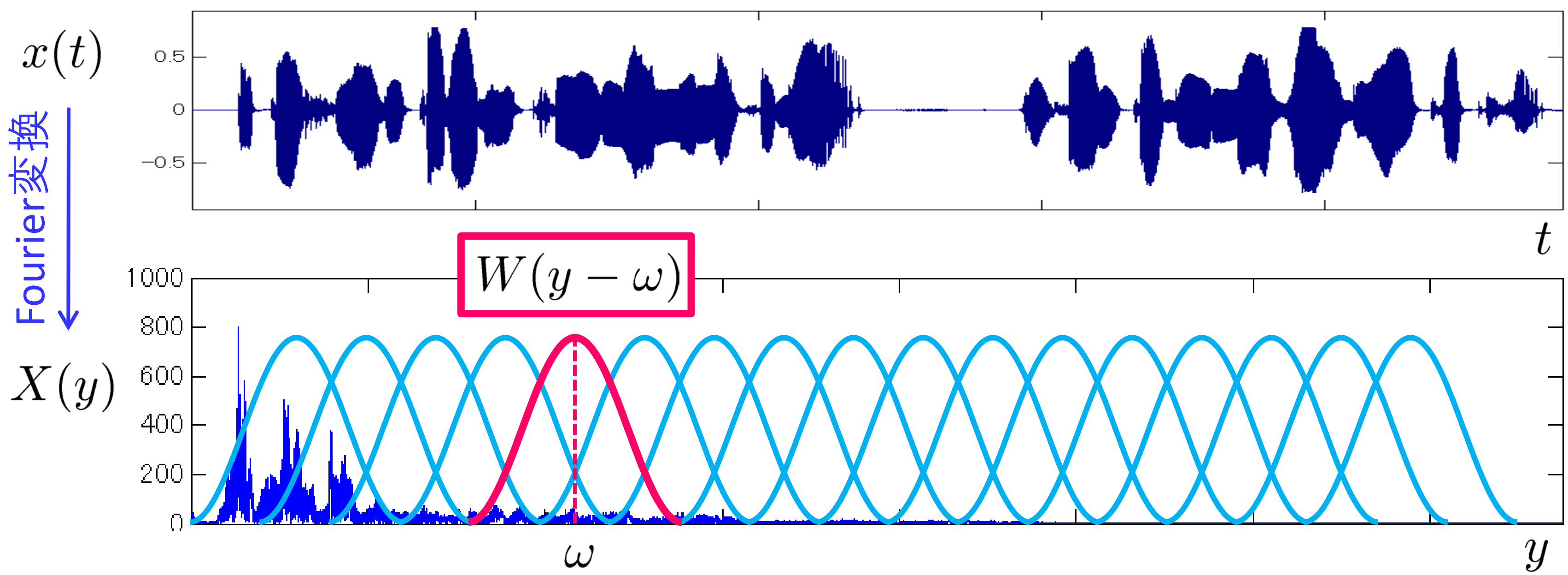
$$\Psi_{\omega,0}(y) = \underline{W}(y-\omega)$$

w のFourier変換

フィルタバンクとしての見方 (2/2)

$$X_{\mathrm{STFT}}(\omega,\tau)=\int_{-\infty}^{\infty}\underline{X(y)W(y-\omega)}e^{jy\tau}\mathrm{d}y$$

$X(y)W(y-\omega)$ の逆Fourier変換



$X_{\mathrm{STFT}}(\omega,\tau)$ は中心周波数が ω のバンドパスフィルタを通過したサブバンド信号と見なせる

時間周波数解析

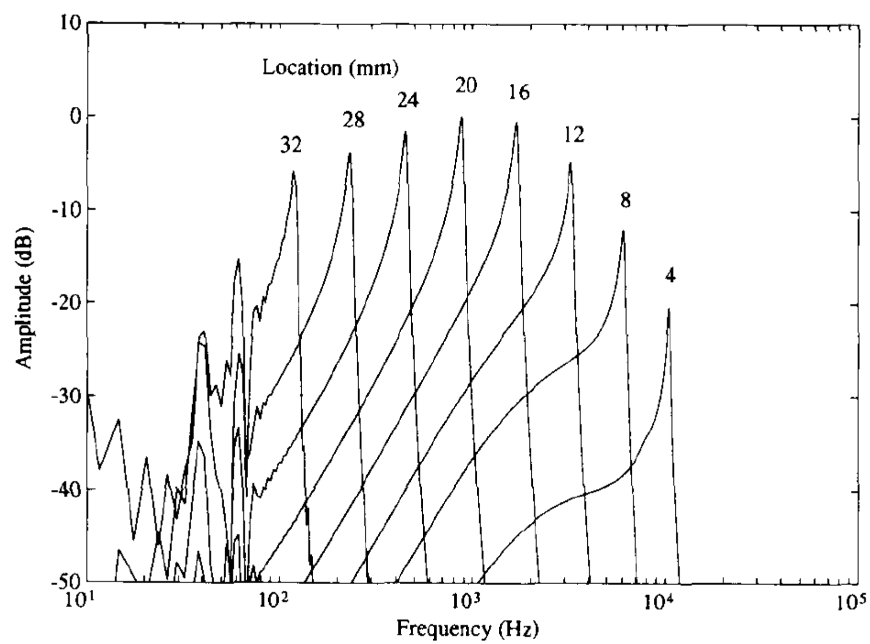
- 動機について
- 短時間Fourier変換 (Short Time Fourier Transform)
  - 定義
  - スペクトログラムとは
  - フィルタバンクとしての見方
- 聴覚フィルタバンク
  - 聴覚システムにおける時間周波数解析
  - 蝸牛モデル
- ウェーブレット変換（定Qフィルタバンク）
  - 定義
  - フィルタバンクとしての見方
- スペクトル漏れ
- スペクトログラムから信号への逆変換

時間周波数解析

- 動機について
- 短時間Fourier変換 (Short Time Fourier Transform)
  - 定義
  - スペクトログラムとは
  - フィルタバンクとしての見方
- 聴覚フィルタバンク
  - 聴覚システムにおける時間周波数解析
  - 蝸牛モデル
- ウェーブレット変換（定Qフィルタバンク）
  - 定義
  - フィルタバンクとしての見方
- スペクトル漏れ
- スペクトログラムから信号への逆変換

聴覚フィルタバンク

- 人間の聴覚システムでは蝸牛と呼ばれる器官で時間周波数解析に相当する処理が行われていると考えられている
  - 蝸牛管の内部は，リンパ液で満たされている
  - 鼓膜，耳小骨を経た振動はリンパを介して蝸牛管内にある基底膜に伝わり，最終的に蝸牛神経を通じて中枢神経に情報が送られる
  - 基底膜は奥にいくほど幅広くかつ柔軟になっており，基部より頂部の方が 曲がりやすく，基部から頂部に至るほどより低い音に対応する固有振動数を持つ
  - 波が基底膜のどの位置まで到達するかで周波数成分が分かる
- 蝸牛モデル [von Békésy1960]
  - 基底膜の各位置における周波数応答は右図のとおり
  - Q値がほぼ等しい



[Irino1993]より抜粋

聴覚フィルタバンク（ムービー）

時間周波数解析

- 動機について
- 短時間Fourier変換 (Short Time Fourier Transform)
  - 定義
  - スペクトログラムとは
  - フィルタバンクとしての見方
- 聴覚フィルタバンク
  - 聴覚システムにおける時間周波数解析
  - 蝸牛モデル
- ウェーブレット変換（定Qフィルタバンク）
  - 定義
  - フィルタバンクとしての見方
- スペクトル漏れ
- スペクトログラムから信号への逆変換

時間周波数解析

- 動機について
- 短時間Fourier変換 (Short Time Fourier Transform)
  - 定義
  - スペクトログラムとは
  - フィルタバンクとしての見方
- 聴覚フィルタバンク
  - 聴覚システムにおける時間周波数解析
  - 蝸牛モデル
- ウェーブレット変換（定Qフィルタバンク）
  - 定義
  - フィルタバンクとしての見方
- スペクトル漏れ
- スペクトログラムから信号への逆変換

ウェーブレット変換（定Qフィルタバンク）

- <u>動機</u>：人間の蝸牛と似た性質をもつ時間周波数解析の方法は？
  - 先に見たとおり，STFTは「定バンド幅フィルタバンク」に相当
  - 等しいQ値のサブバンドフィルタからなるフィルタバンクが考えられないか？

→ ウェーブレット変換（定Qフィルタバンク）

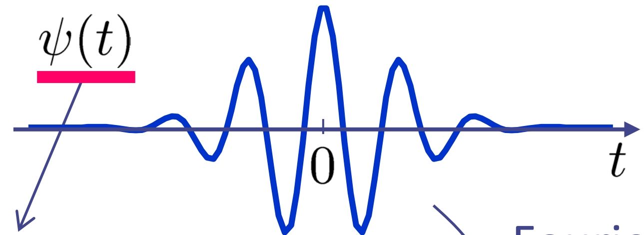
ウェーブレット変換（定Qフィルタバンク）

■定義：信号と「ウェーブレット」（小さい波）との内積

$$X_{\mathrm{wavelet}}(\alpha,\tau) = \langle x(t), \underline{\psi_{\alpha,\tau}(t)} \rangle_{t\in\mathbb{R}} = \int_{-\infty}^{\infty} x(t)\psi^*_{\alpha,\tau}(t)\mathrm{d}t$$

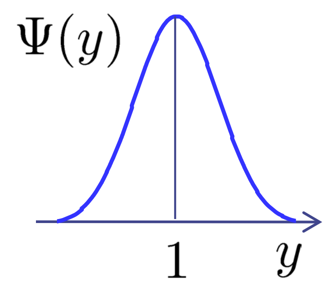
$$\underline{\psi_{\alpha,t}(t)} := \frac{1}{\alpha}\underline{\psi}\left(\frac{t-\tau}{\alpha}\right)$$

基底関数



アナライジングウェーブレット
（中心周波数が1の波）

Fourier
変換



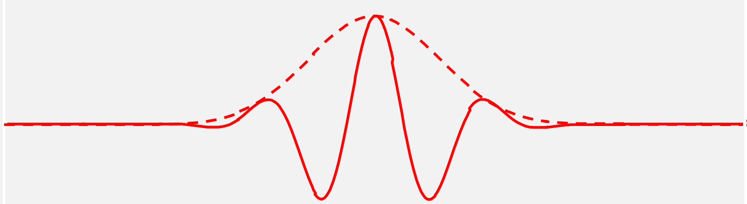
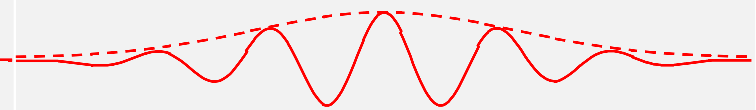
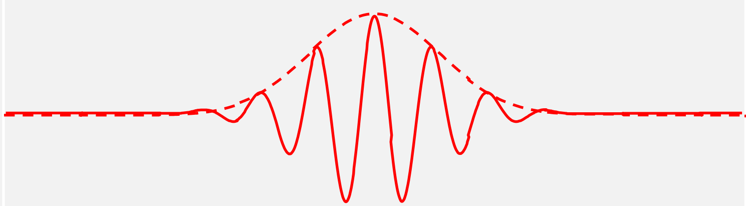
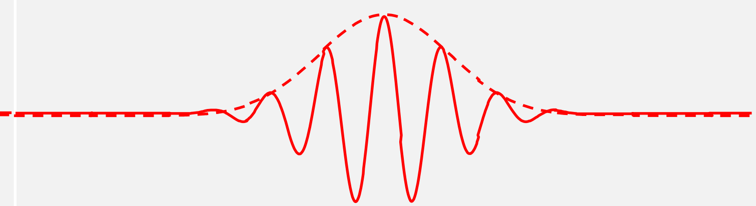
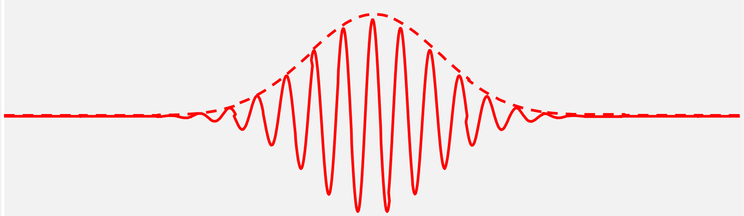
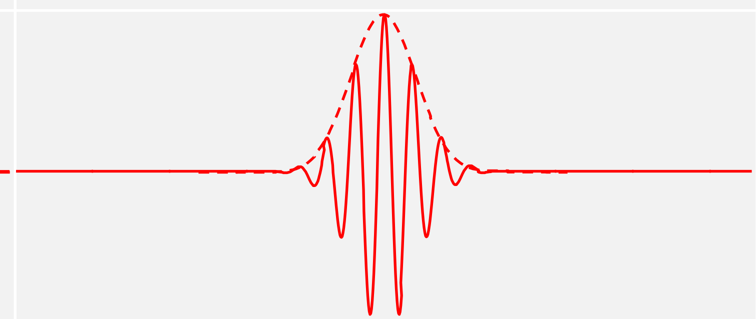
$\underline{\psi_{\alpha,\tau}}$ はアナライジングウェーブレットを時間方向に α 倍縮めて，τ だけシフトしたもの

（時刻 τ に局在する周期 α の小さい波）

$X_{\mathrm{wavelet}}(\alpha,\tau)$ は $x(t)$ に含まれる，
時刻 τ 周辺における周期 α の成分に相当

STFTとの違い

■周波数ごとの基底関数 ψ の比較

周波数	STFT	ウェーブレット変換
低		
↕		
高		

フィルタバンクとしての見方（本当に「定Q」なのか？）

$$X_{\mathrm{wavelet}}(\alpha,\tau) = \langle x(t), \psi_{\alpha,\tau}(t)\rangle_{t\in\mathbb{R}}$$

$$= \langle X(y), \underbrace{\Psi_{\alpha,\tau}(y)}_{\psi_{\alpha,\tau}\ \text{のFourier変換}}\rangle_{y\in\mathbb{R}}$$

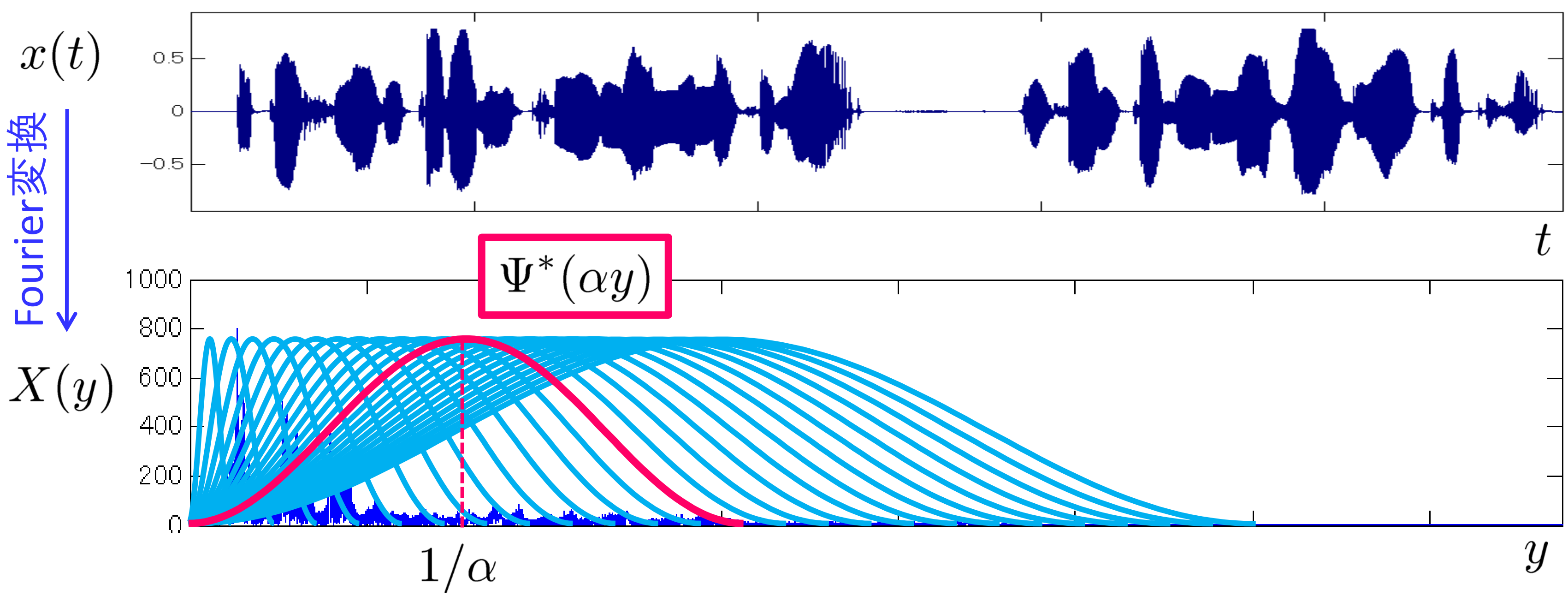
$$= \int_{-\infty}^{\infty} X(y)\Psi^*(\alpha y)e^{jy\tau}\mathrm{d}y$$

$X(y)\Psi^*(\alpha y)$ の逆Fourier変換

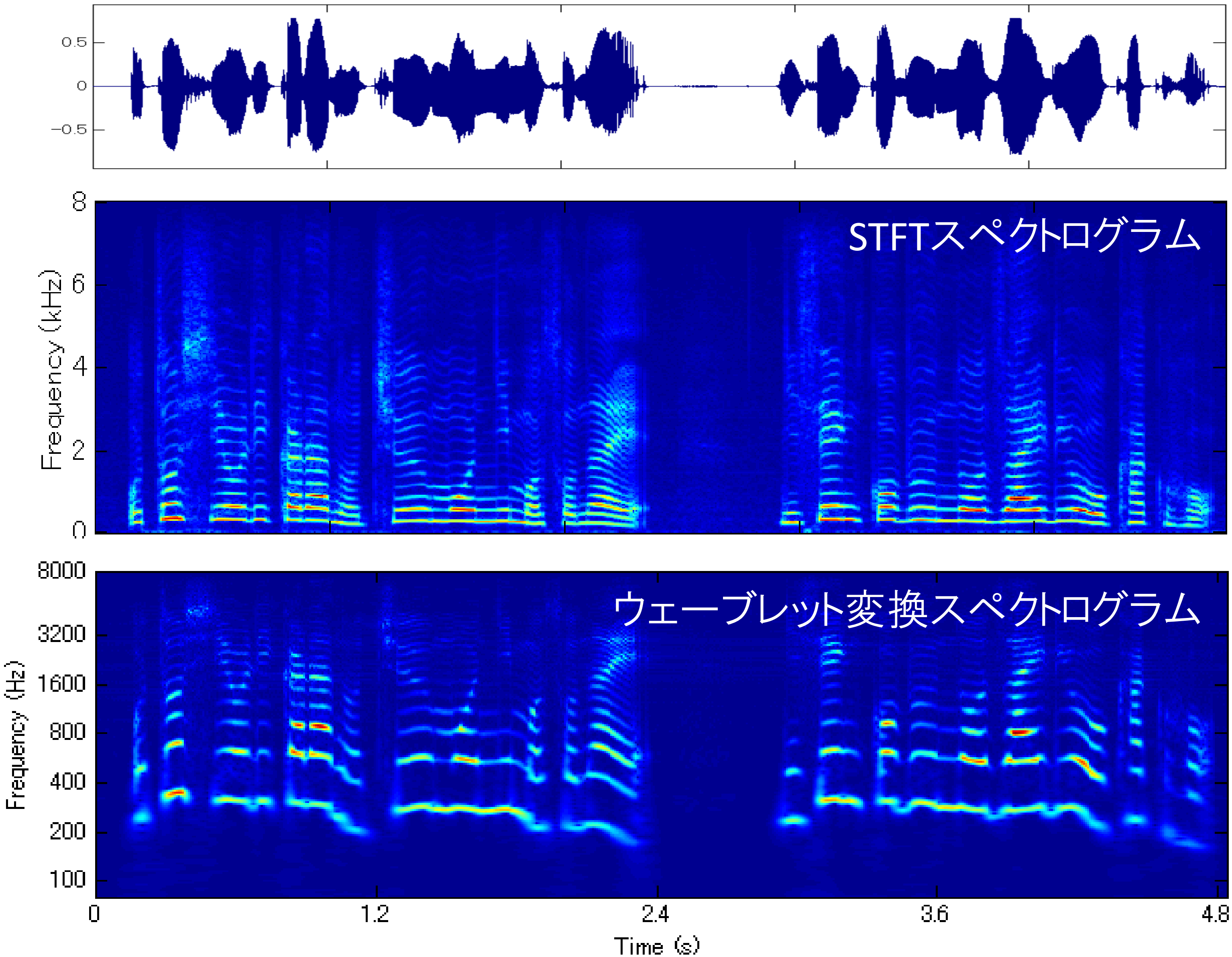
一般化Parsevalの定理：
時間領域の内積は
周波数領域の内積と等しい

$\psi_{\alpha,t}(t) = \frac{1}{\alpha}\psi\left(\frac{t-\tau}{\alpha}\right)$ より

$$\Psi_{\alpha,\tau}(y) = \underbrace{\Psi}_{\psi\ \text{のFourier変換}}(\alpha y)e^{-jy\tau}$$



STFTとウェーブレット変換によるスペクトログラムの比較



時間周波数解析

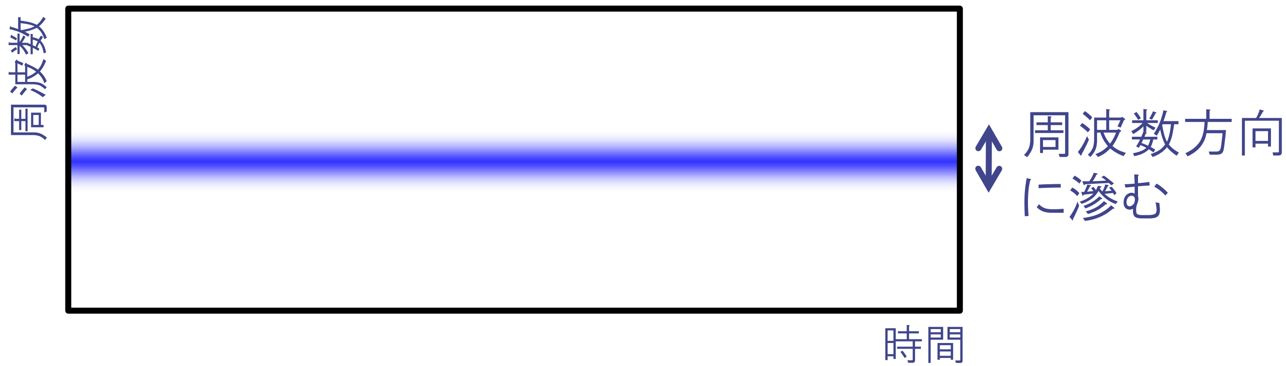
- 動機について
- 短時間Fourier変換 (Short Time Fourier Transform)
  - 定義
  - スペクトログラムとは
  - フィルタバンクとしての見方
- 聴覚フィルタバンク
  - 聴覚システムにおける時間周波数解析
  - 蝸牛モデル
- ウェーブレット変換（定Qフィルタバンク）
  - 定義
  - フィルタバンクとしての見方
- スペクトル漏れ
- スペクトログラムから信号への逆変換

時間周波数解析

- 動機について
- 短時間Fourier変換 (Short Time Fourier Transform)
  - 定義
  - スペクトログラムとは
  - フィルタバンクとしての見方
- 聴覚フィルタバンク
  - 聴覚システムにおける時間周波数解析
  - 蝸牛モデル
- ウェーブレット変換（定Qフィルタバンク）
  - 定義
  - フィルタバンクとしての見方
- スペクトル漏れ
- スペクトログラムから信号への逆変換

スペクトル漏れ

- たとえ信号が無限長の正弦波であっても，STFTやウェーブレット変換のスペクトログラムは周波数方向にエネルギーが拡散される



- このスペクトルの広がりがGauss分布の形になる分析条件は？
  - 音源分離や多重ピッチ推定のための効果的な手法として，スペクトログラムをGauss分布を用いてモデル化するアプローチが提案されている [Kameoka2007]
  - スペクトルの広がりがGauss分布の形になる分析条件を明らかにすれば，上述のモデルで置いた仮定と理想的に一致したスペクトログラムを得ることができる
- STFTの場合は明らか　→　窓関数をGauss窓にする（Gabor変換）
  ウェーブレット変換の場合は？

ウェーブレット変換スペクトルがGauss分布形になる条件

$x(t) = we^{j\omega_0 t}$ として，そのウェーブレット変換スペクトログラム $|X_{\mathrm{wavelet}}(e^{-u},\tau)|^2$ が対数周波数 u 方向にGauss分布の形になるようにするためには，アナライジングウェーブレットをどのような関数にしたら良いか？ [Kameoka2007]

$$X_{\mathrm{wavelet}}(\alpha,\tau) = \langle x(t), \psi_{\alpha,\tau}(t)\rangle_{t\in\mathbb{R}}$$

$$= \int_{-\infty}^{\infty} X(y)\Psi^*(\alpha y)e^{jy\tau}\mathrm{d}y$$

$$X(y) = \sqrt{2\pi}w\delta(y-\omega_0)$$

$$= \sqrt{2\pi}w\int_{-\infty}^{\infty} \delta(y-\omega_0)\Psi^*(\alpha y)e^{jy\tau}\mathrm{d}y$$

$$= \sqrt{2\pi}w\Psi^*(\alpha\omega_0)e^{j\omega_0\tau}$$

$$u = \log\frac{1}{\alpha} \to \alpha = e^{-u}$$

$$X_{\mathrm{wavelet}}(e^{-u},\tau) = \sqrt{2\pi}w\Psi^*(e^{-u+\log\omega_0})e^{j\omega_0\tau}$$

$$= \sqrt{2\pi}we^{-\frac{(u-\log\omega_0)^2}{4\sigma^2}}e^{j\omega_0\tau}$$

$$|X_{\mathrm{wavelet}}(e^{-u},\tau)|^2 = 2\pi w^2 e^{-\frac{(u-\log\omega_0)^2}{2\sigma^2}}$$

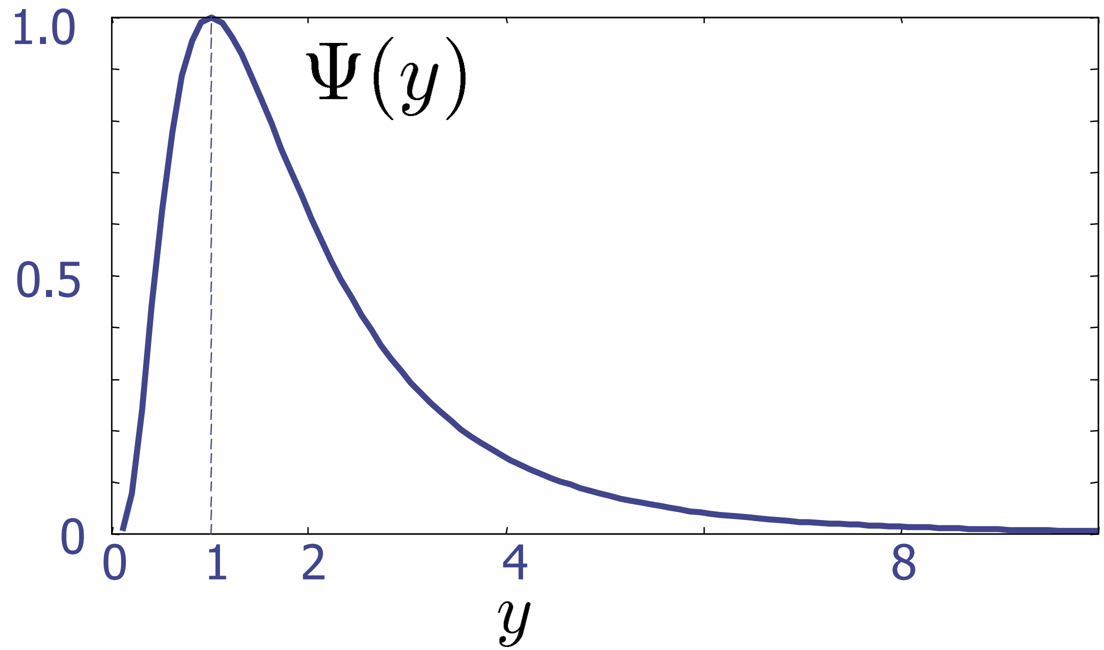
$$\Psi(y) = \begin{cases} e^{-\frac{(\log y)^2}{4\sigma^2}} & y>0 \\ 0 & y\le 0 \end{cases}$$

「定Q 対数正規分布型フィルタバンク」[Kameoka2007]

$$x(t) = we^{j\omega_0 t}$$

$$\Psi(y) = \begin{cases} e^{-\frac{(\log y)^2}{4\sigma^2}} & y > 0 \\ 0 & y \leq 0 \end{cases} \longrightarrow |X_{\mathrm{wavelet}}(e^{-u}, \tau)|^2 = 2\pi w^2 e^{-\frac{(u - \log \omega_0)^2}{2\sigma^2}}$$

アナライジングウェーブレットの周波数応答が対数分布形



時間周波数解析

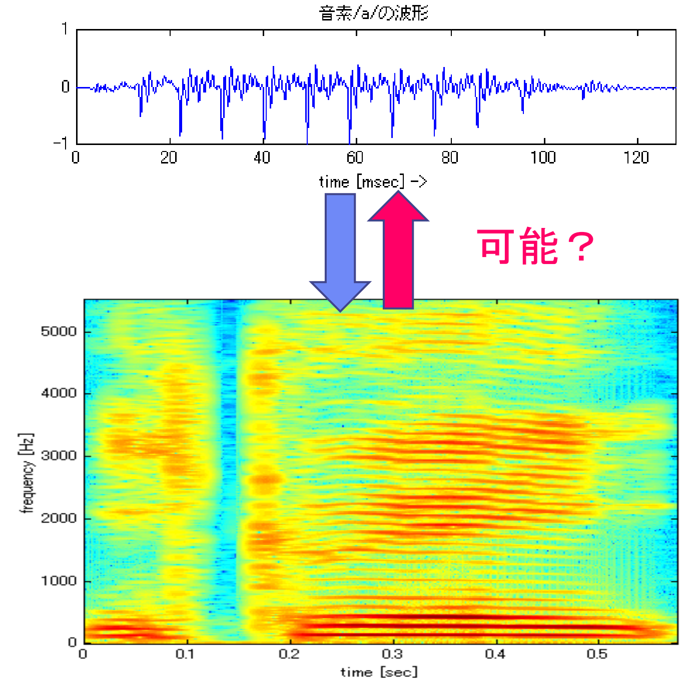
- 動機について
- 短時間Fourier変換 (Short Time Fourier Transform)
  - 定義
  - スペクトログラムとは
  - フィルタバンクとしての見方
- 聴覚フィルタバンク
  - 聴覚システムにおける時間周波数解析
  - 蝸牛モデル
- ウェーブレット変換（定Qフィルタバンク）
  - 定義
  - フィルタバンクとしての見方
- スペクトル漏れ
- スペクトログラムから信号への逆変換

時間周波数解析

- 動機について
- 短時間Fourier変換 (Short Time Fourier Transform)
  - 定義
  - スペクトログラムとは
  - フィルタバンクとしての見方
- 聴覚フィルタバンク
  - 聴覚システムにおける時間周波数解析
  - 蝸牛モデル
- ウェーブレット変換(定Qフィルタバンク)
  - 定義
  - フィルタバンクとしての見方
- スペクトル漏れ
- スペクトログラムから信号への逆変換

スペクトログラムから信号への逆変換

- 信号→スペクトログラム
  - 時間周波数解析
    - 短時間Fourier変換
    - ウェーブレット変換
  - 信号の特徴が視覚化できて便利
  - 位相を捨てるので不可逆変換



- スペクトログラム→信号
  - 信号の加工をスペクトログラムの加工を通して行えて便利
  - 「信号→スペクトログラム」は不可逆変換と言ったばかり！
  - これを行うための逆変換

反復STFT [Griffin1984]

スペクトログラム（各時間周波数のパワー）のみが既知

(Step0) 各時間周波数における位相を適当に設定
- 仮の複素スペクトル時系列の獲得

(Step1) 逆STFT
- 信号への変換

(Step2) STFT
- 複素スペクトル時系列の獲得

(Step3) パワーの置換
- 位相はそのまま、パワーだけ既知のものに置き換え

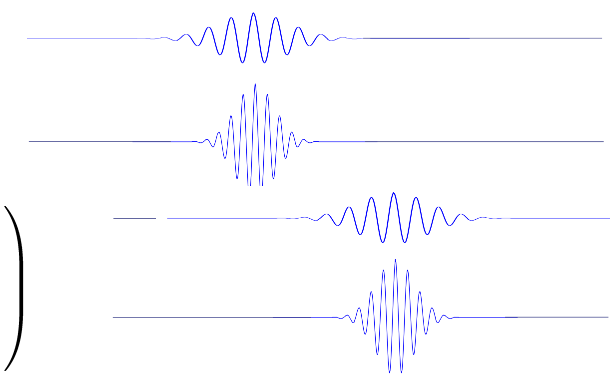
時間周波数分解の一般表現

- 離散時間信号 $\boldsymbol{x} = (x_1, \ldots, x_K)^{\mathrm{T}}$
- 時間周波数解析
  - ある時刻の周波数成分は，ある特定の基底関数との**内積**によって計測できる

1フレーム目の時刻を中心に局在する，周波数ごとの「ウェーブレット」基底

$$\boldsymbol{f} = \boldsymbol{W}\boldsymbol{x}$$

$$\begin{array}{l} \text{1フレーム目} \\ \\ \text{2フレーム目} \\ \\ \\ \\ \end{array} \begin{pmatrix} f_1 \\ \vdots \\ f_{N_1} \\ \hline f_{N_1+1} \\ \vdots \\ f_{N_1+N_2} \\ \hline \vdots \\ f_N \end{pmatrix} = \begin{pmatrix} w_{1,1} & \cdots & w_{1,K} \\ \vdots & & \vdots \\ w_{N_1,1} & \cdots & w_{N_1,K} \\ \hline w_{N_1+1} & \cdots & w_{N_1+1,K} \\ \vdots & & \vdots \\ w_{N_1+N_2} & \cdots & w_{N_1+N_2,K} \\ \hline \vdots & & \vdots \\ w_N & \cdots & w_{N,K} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_K \end{pmatrix}$$



$w : \mathbb{R}^K \to \mathbb{C}^N$

フレーム数×周波数bin数

$N > K$

$\mathbf{W}$は縦長行列

スペクトログラムから信号を復元する問題

- 時間周波数平面上の各点における振幅が既知

$$\boldsymbol{c} = (c_1, \ldots, c_N)^{\mathrm{T}}$$

- $\boldsymbol{f} = \boldsymbol{W}\boldsymbol{x},\ \mathrm{abs}(\boldsymbol{f}) = \boldsymbol{c}$ となる $\boldsymbol{x}$ を見つけたい

→解が存在するとは限らない

ただし, $\mathrm{abs}(\boldsymbol{f}) \equiv (|f_1|, \cdots, |f_N|)^{\mathrm{T}}$

問題の定式化とその解法

- $\boldsymbol{f} \simeq \boldsymbol{W}\boldsymbol{x},\ \mathrm{abs}(\boldsymbol{f}) = \boldsymbol{c}$ となる $\boldsymbol{x}$ を見つけたい

$$\begin{aligned} &\text{minimize} \quad \left\| \boldsymbol{f} - \boldsymbol{W}\boldsymbol{x} \right\|_2^2 \\ &\text{subject to} \quad \mathrm{abs}(\boldsymbol{f})| = \boldsymbol{c} \end{aligned}$$

(Step0) $\boldsymbol{f}$ を初期設定

(Step1) $\boldsymbol{x} \leftarrow \underset{\boldsymbol{x}}{\operatorname{argmin}} \left\| \boldsymbol{f} - \boldsymbol{W}\boldsymbol{x} \right\|_2^2$

$$= \left(\boldsymbol{W}^H \boldsymbol{W}\right)^{-1} \boldsymbol{W}^H \boldsymbol{f}^{(t)}$$ （逆STFT，逆ウェーブレット変換）

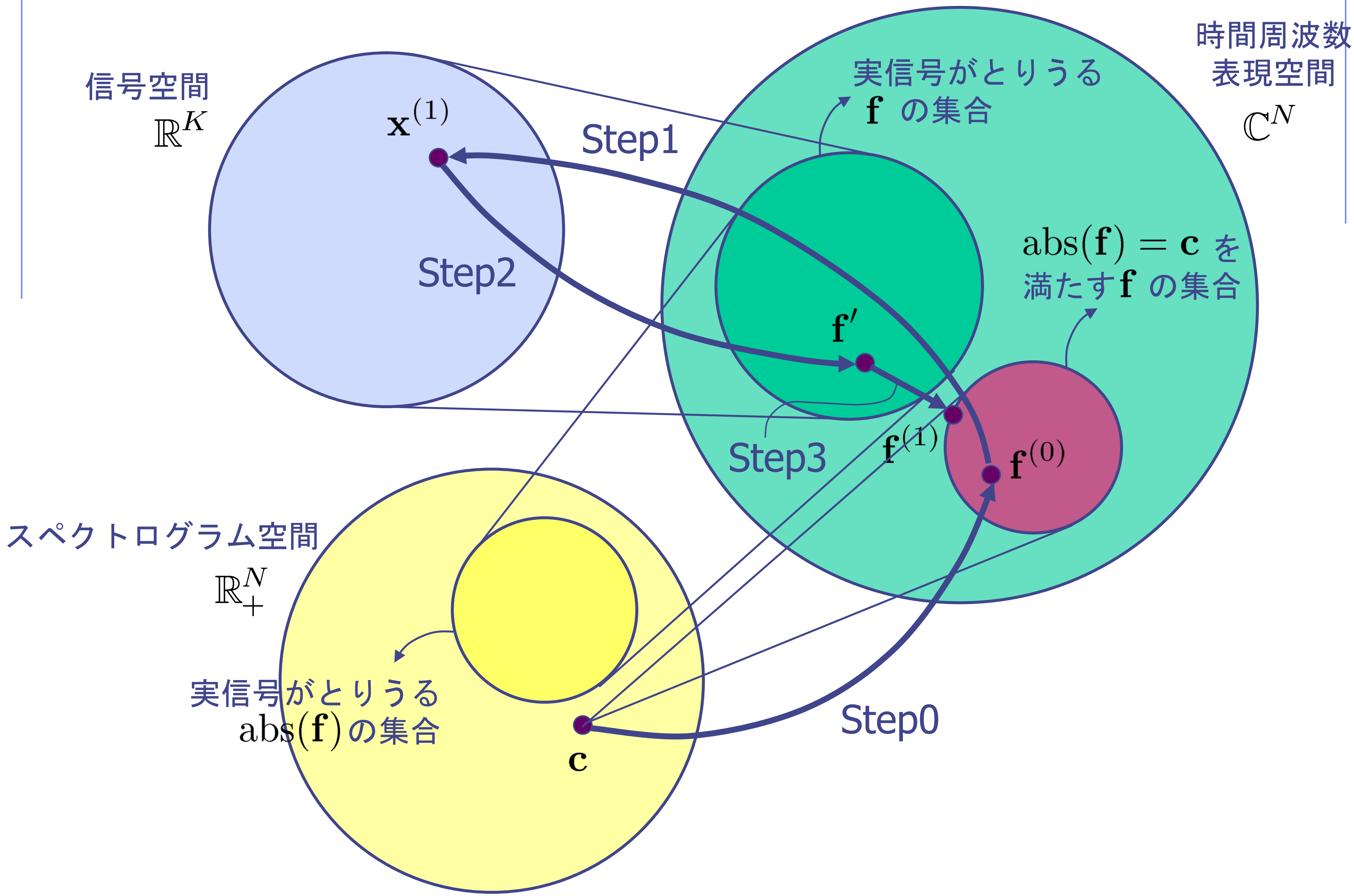
(Step2) $\boldsymbol{f} \leftarrow \underset{\boldsymbol{f}}{\operatorname{argmin}} \left\| \boldsymbol{f} - \boldsymbol{W}\boldsymbol{x} \right\|_2^2$

$$= \boldsymbol{W}\boldsymbol{x}$$ （STFT，ウェーブレット変換）

(Step3) $\boldsymbol{f} \leftarrow \boldsymbol{c} \odot \mathrm{phase}(\boldsymbol{f})$ （パワーの置換）

Step1へ

反復STFTの図解



信号空間
$\mathbb{R}^K$
$\mathbf{x}^{(1)}$
Step1
Step2
時間周波数
表現空間
$\mathbb{C}^N$
実信号がとりうる
$\mathbf{f}$ の集合
$\mathrm{abs}(\mathbf{f}) = \mathbf{c}$ を
満たす$\mathbf{f}$ の集合
$\mathbf{f}'$
Step3
$\mathbf{f}^{(1)}$
$\mathbf{f}^{(0)}$
スペクトログラム空間
$\mathbb{R}_+^N$
実信号がとりうる
$\mathrm{abs}(\mathbf{f})$ の集合
$\mathbf{c}$
Step0

レポート課題の対象論文

- D.W. Griffin and J.S. Lim, "Signal estimation from modified short-time Fourier transform," IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP-32, No. 2, pp. 236-243, 1984.
- T. Irino and R.D. Patterson, "A time-domain, level-dependent auditory filter: The gammachirp", The Journal of the Acoustic Society of America, Vol. 101, pp. 412–419, 1997.