東京大学工学部 4年生 夏学期 [03-501130]

応用音響学
第2回 (4/15)

亀岡弘和

東京大学大学院情報理工学系研究科
システム情報学専攻
kameoka@hil.t.u-tokyo.ac.jp

講義スケジュール

前半（亀岡担当）

- 4/08: 第1回
- 4/15: 第2回
- 4/22: 第3回
- 4/29: 休日
- 5/01(木): 第4回
- 5/06: 休日
- 5/13: 第5回
- 5/20: 第6回
- 5/27: 第7回

後半（牧野担当）

- 6/03: 第8回
- 6/10: 第9回
- 6/17: 第10回
- 6/24: 第11回
- 7/01: 第12回
- 7/08: 第13回
- 7/15: 第14回
- 7/22: 学期末試験

音響学の研究分野とトピック

- 日本音響学会における研究分野
  - 音声
    - A: 音声認識，対話システム
    - B: 音声分析，合成，符号化
  - 電気(応用)音響
    - マイクロホンアレイ，スピーカ音声強調，音源分離
  - 音楽音響
    - 楽器音響，音楽情報処理
  - 聴覚
    - 聴覚心理，聴覚情報処理
  - 超音波
  - 騒音・振動
  - 建築音響

音声: 線形予測分析　隠れマルコフモデル　混合正規分布モデル　カルマンフィルタ

電気(応用)音響: 独立成分分析　ウィーナーフィルタ　カルマンフィルタ　非負値行列因子分解

音楽音響: 非負値行列因子分解　隠れマルコフモデル

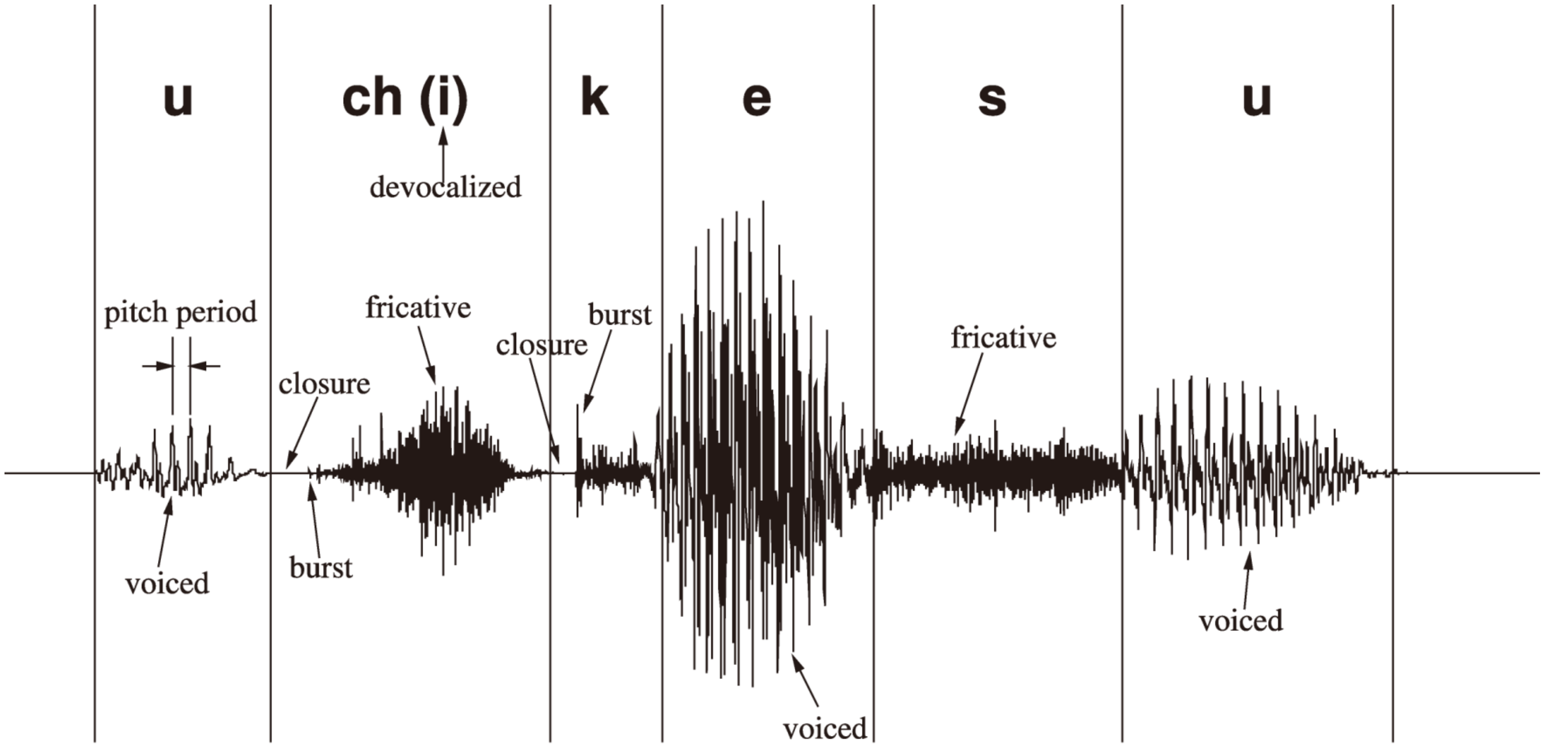
↑各分野の基礎トピック

本日の話題

- サンプリング（標本化）の復習
- 時間周波数解析（短時間スペクトル分析）
  - 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 (<u>S</u>hort <u>T</u>ime <u>F</u>ourier <u>T</u>ransform)
  - ウェーブレット変換（定Qフィルタバンク）

音声波形

- 「打ち消す」の音声波形



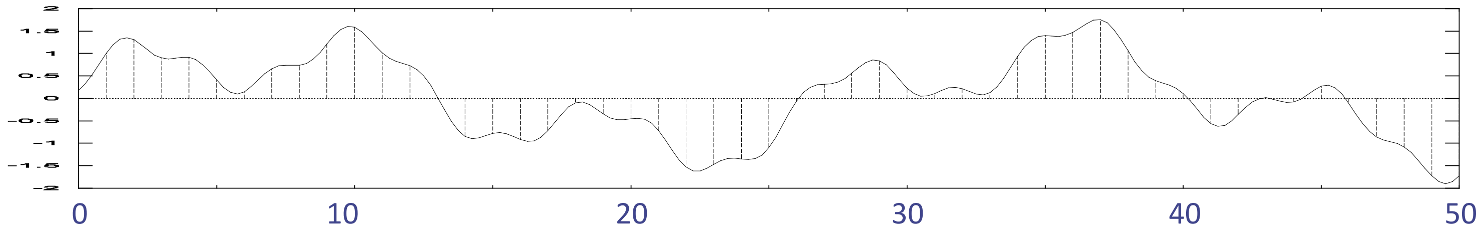
- 「ち」は無声化して母音の/i/が脱落
- 有声音(ここでは/u/, /e/, /u/)の波形は局所的に周期的
  →これがピッチ(声の高さ)として感じられる

音声のディジタル化のプロセス

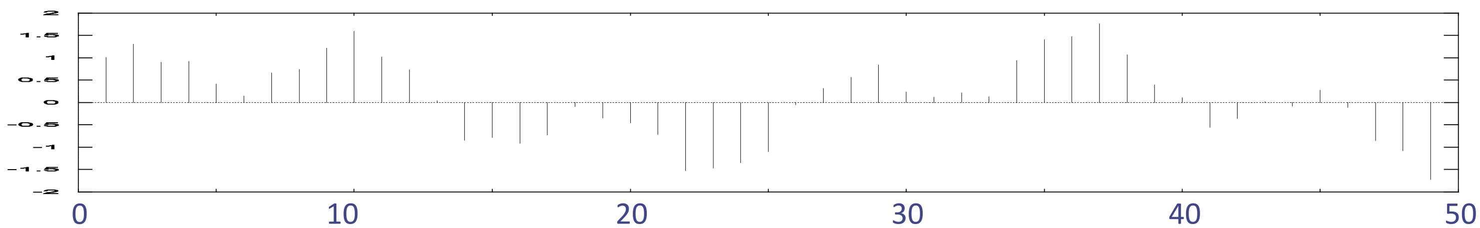
- 音の電気変換（マイクロフォン） (transducer)
- 増幅 (amplification)
- A/D変換 (A-to-D conversion)
  - フィルタリング (filtering)
  - サンプリング (sampling)
  - 量子化 (quantization)
- ディジタル値の取り込み

信号のサンプリング（標本化）と復元

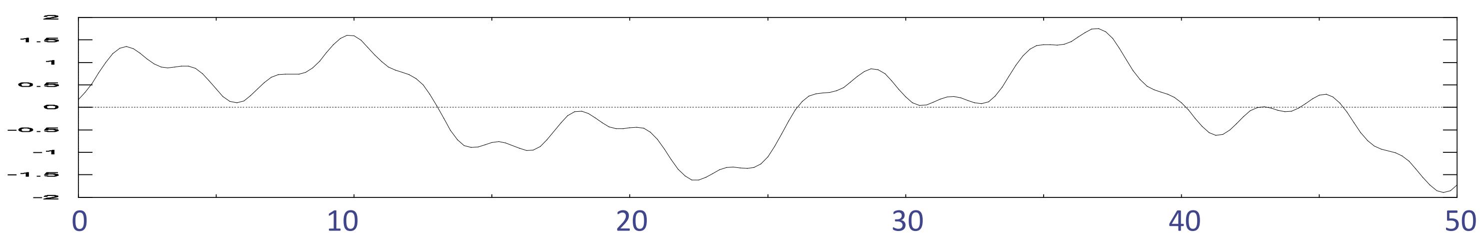
■連続時間信号



■サンプリング（標本化）



■信号復元



標本値系列からの信号復元

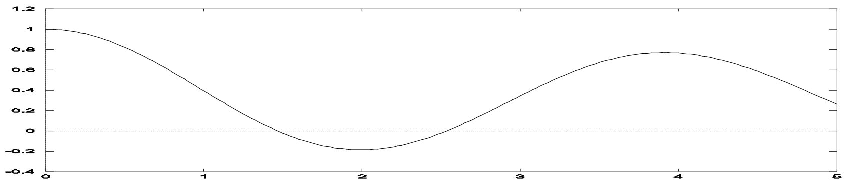
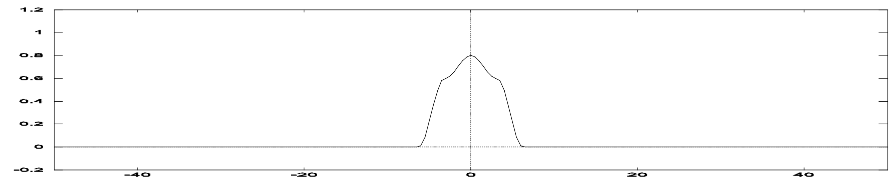
- 正確に復元できる(原情報を保存できる)条件は？
- 標本値系列の多義性
- 直感的には
  - 正弦波1周期あたり標本点が2点必要

標本化定理

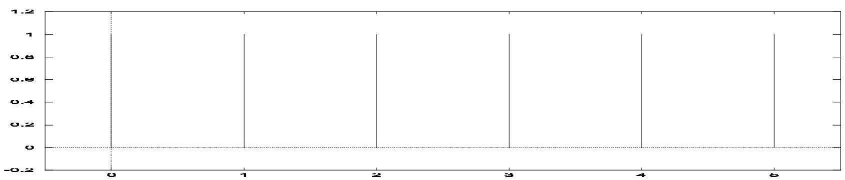
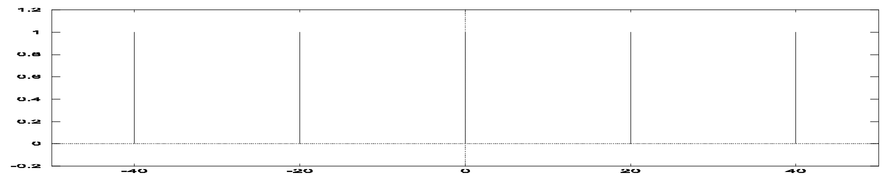
帯域制限信号に対し、カットオフ周波数の2倍以上で標本化すれば、標本値系列から元の信号が完全に復元される。

サンプリング定理(標本化定理)

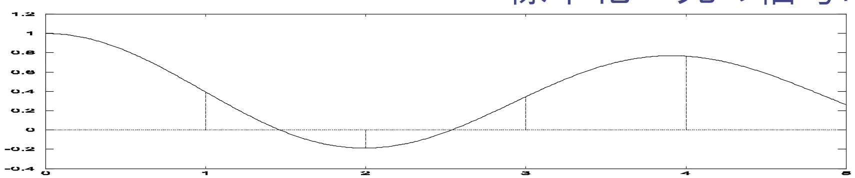
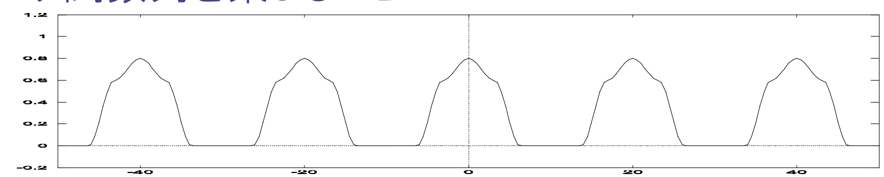
信号波形(時間領域)　　　スペクトル(周波数領域)

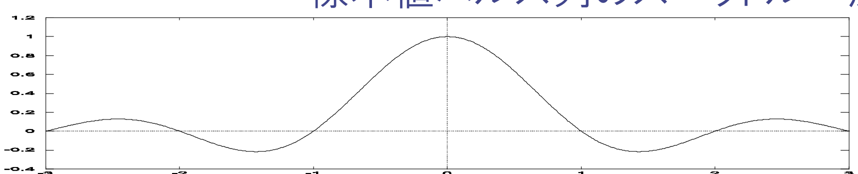
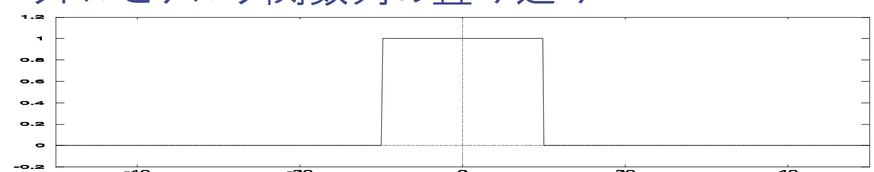
元の信号とスペクトル(帯域が制限されている)

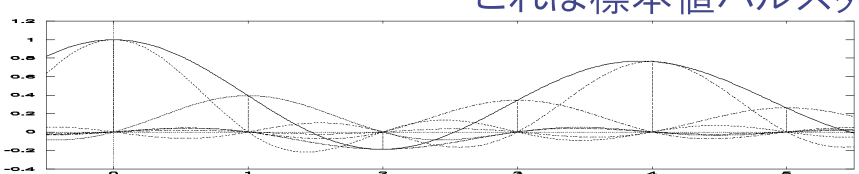
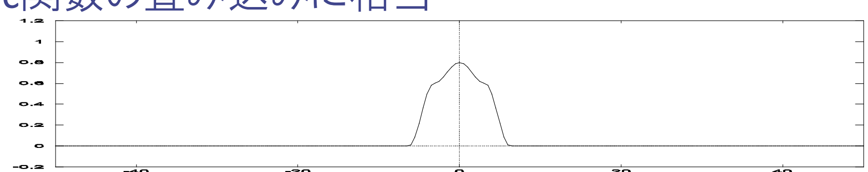
標本化＝元の信号にデルタ関数列を乗じること

標本値パルス列のスペクトル＝原スペクトルとデルタ関数列の畳み込み

スペクトルに矩形関数を乗じたもの＝原スペクトルと同じスペクトル
これは標本値パルス列とsinc関数の畳み込みに相当

時間周波数解析（短時間スペクトル分析）

- 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

時間周波数解析（短時間スペクトル分析）

- 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

時間周波数解析の動機

- Fourier変換：

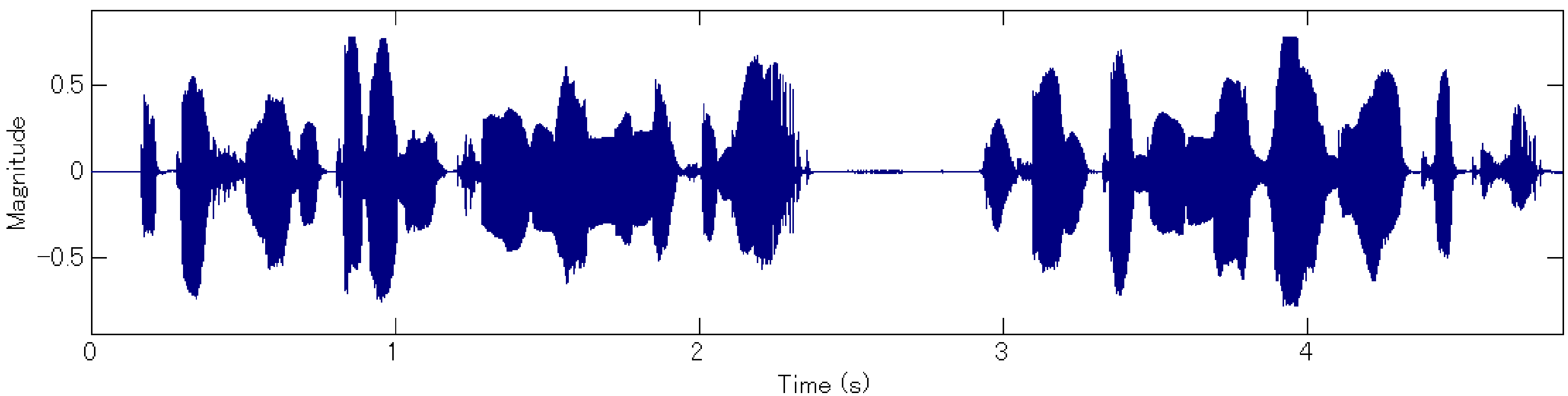
周波数ωの複素正弦波との内積

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

  - $|X(\omega)|$：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

- 音声などの音響信号は非定常（非定常だから情報が運べる）



  - 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（短時間スペクトル分析）

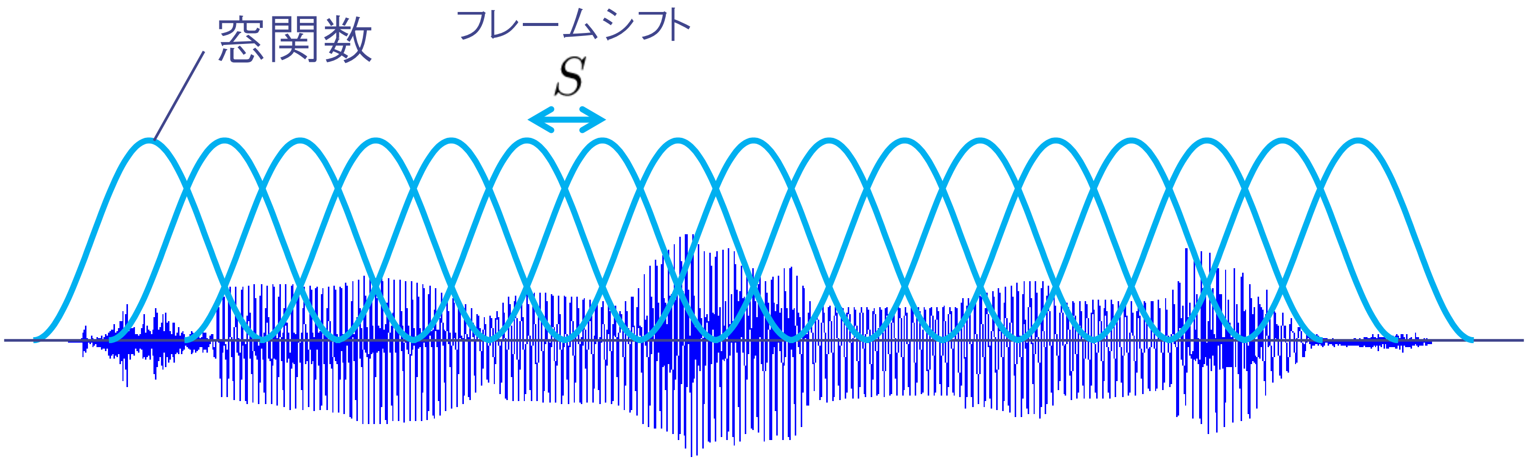
- 動機について
- 短時間Fourier変換 (ShortTimeFourierTransform)
  - 定義
  - スペクトログラムとは
  - フィルタバンクとしての見方
- 聴覚フィルタバンク
  - 聴覚システムにおける時間周波数解析
  - 蝸牛モデル
- ウェーブレット変換（定Qフィルタバンク）
  - 定義
  - フィルタバンクとしての見方

時間周波数解析（短時間スペクトル分析）

- 動機について
- 短時間Fourier変換 (<u>S</u>hort<u>T</u>ime<u>F</u>ourier<u>T</u>ransform)
  - 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

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



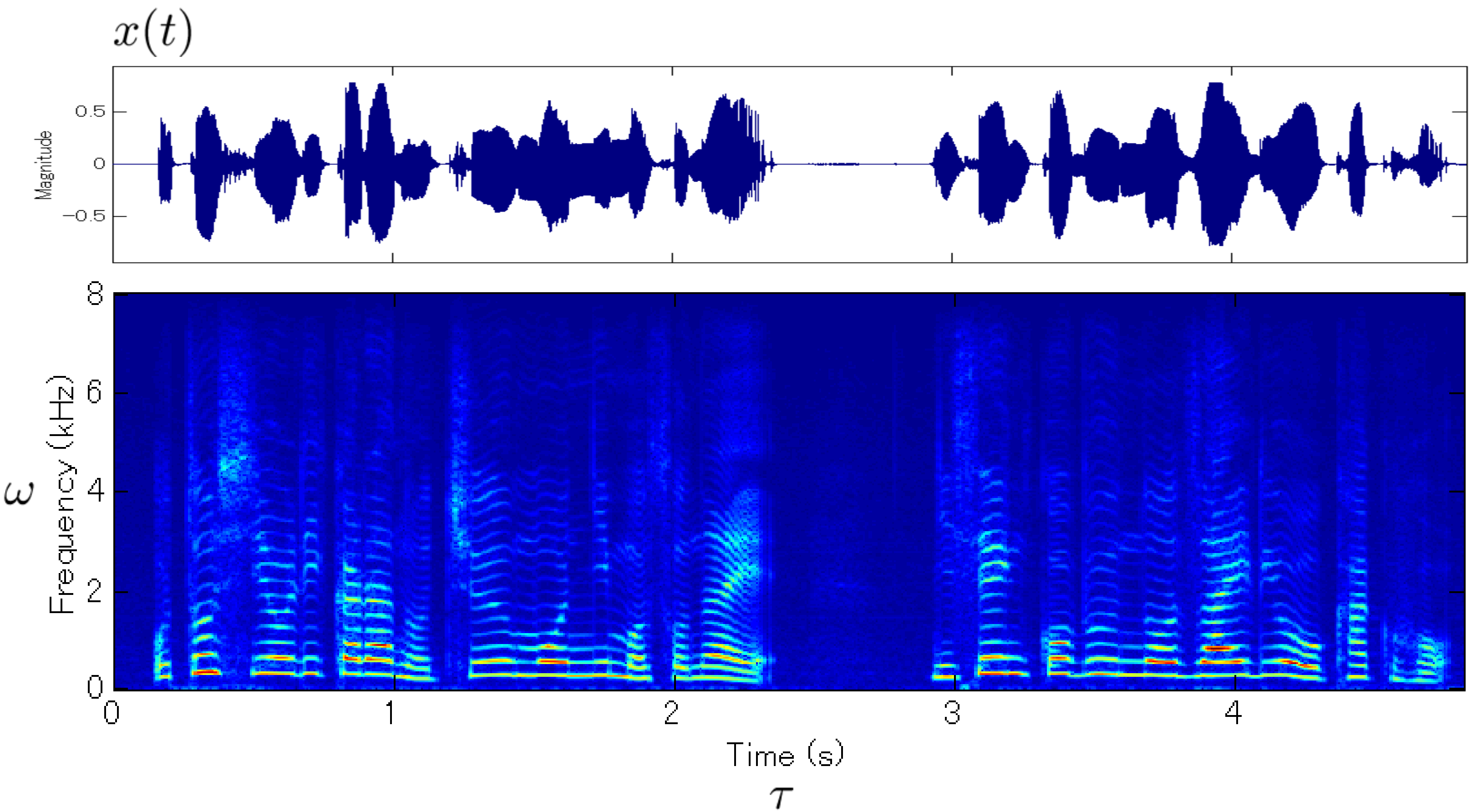
- 式で書くと・・・

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)

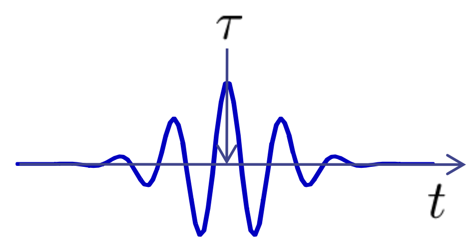
$$X_{\mathrm{STFT}}(\omega,\tau)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}w(t)x(t+\tau)e^{-j\omega t}\mathrm{d}t$$

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

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

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

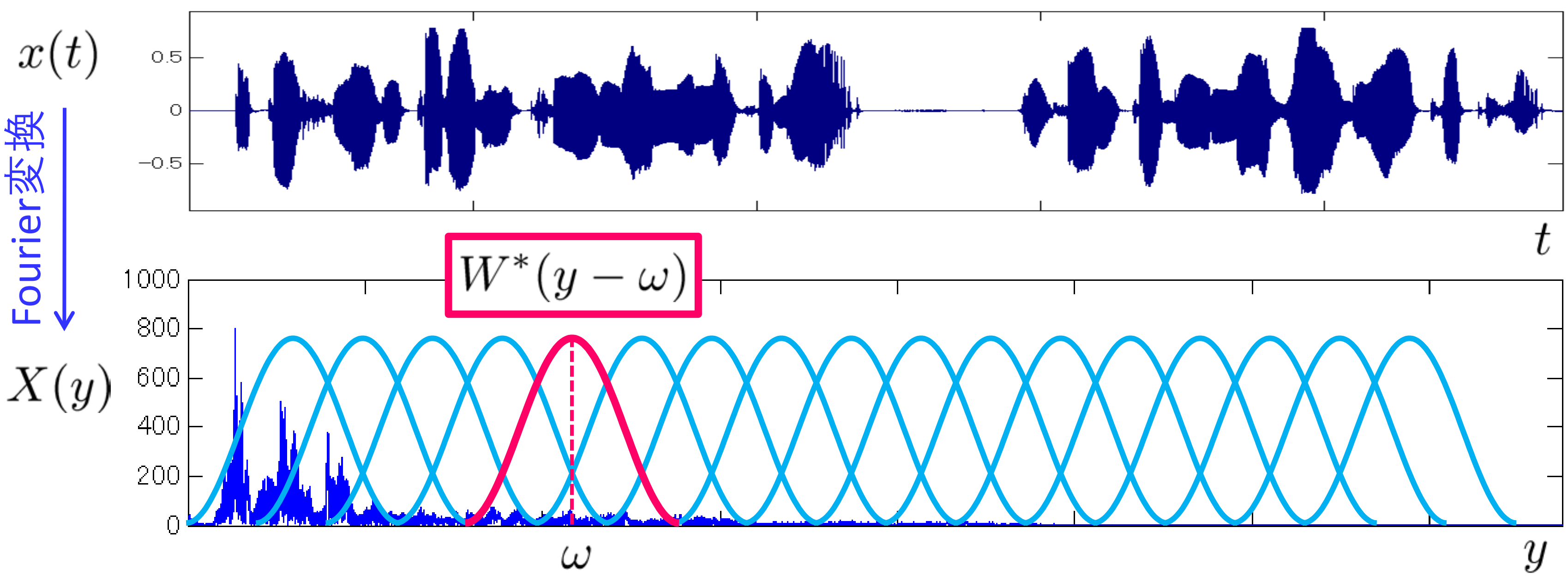
$$\because\ \Psi_{\omega,\tau}(y)=\Psi_{\omega,0}(y)e^{-jy\tau}$$
$$\Psi_{\omega,0}(y)=\underline{W}(y-\omega)$$

w のFourier変換

フィルタバンクとしての見方 (2/2)

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

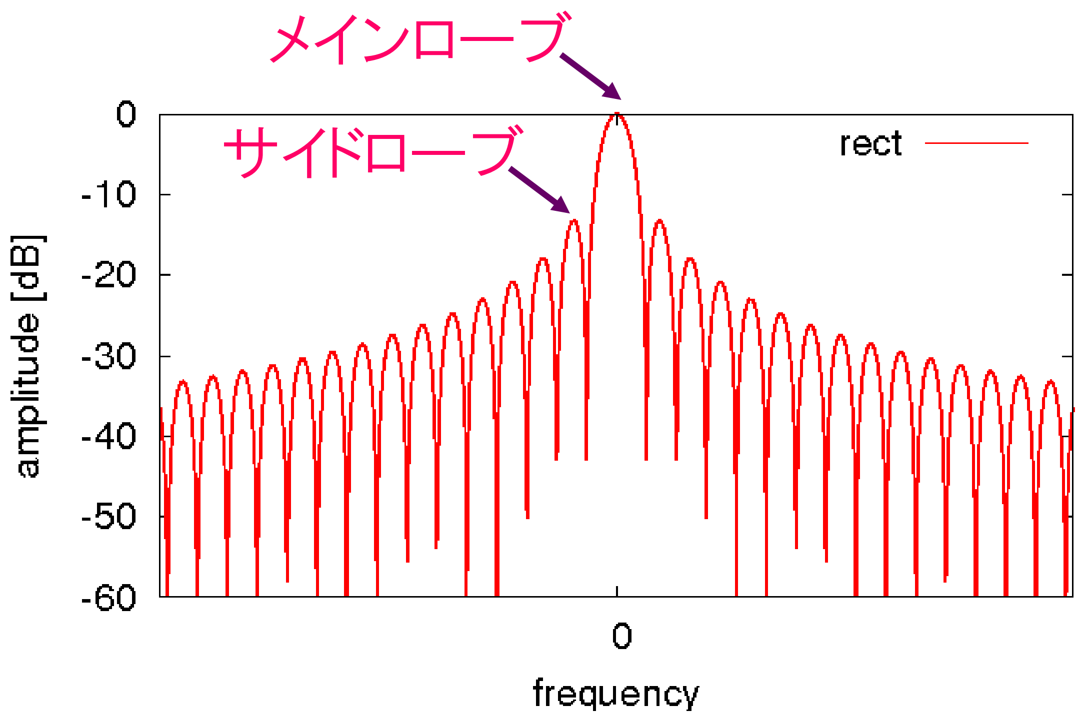
$X(y)W^*(y-\omega)$ の逆Fourier変換



$X_{\mathrm{STFT}}(\omega,\tau)$ 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

様々な窓関数

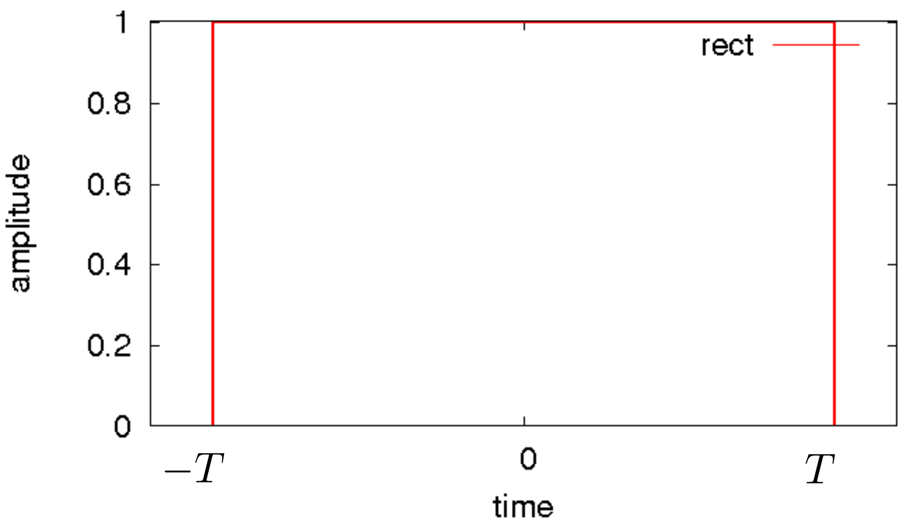
- 窓関数の重要な特性
  - メインローブの幅（周波数分解能）：狭いほどよい
  - サイドローブの大きさ（ダイナミックレンジ）：小さいほどよい
- 代表的な窓関数
  - 矩形窓
  - 三角窓（Bartlett窓）
  - Hanning窓
  - Hamming窓
  - Blackmann窓

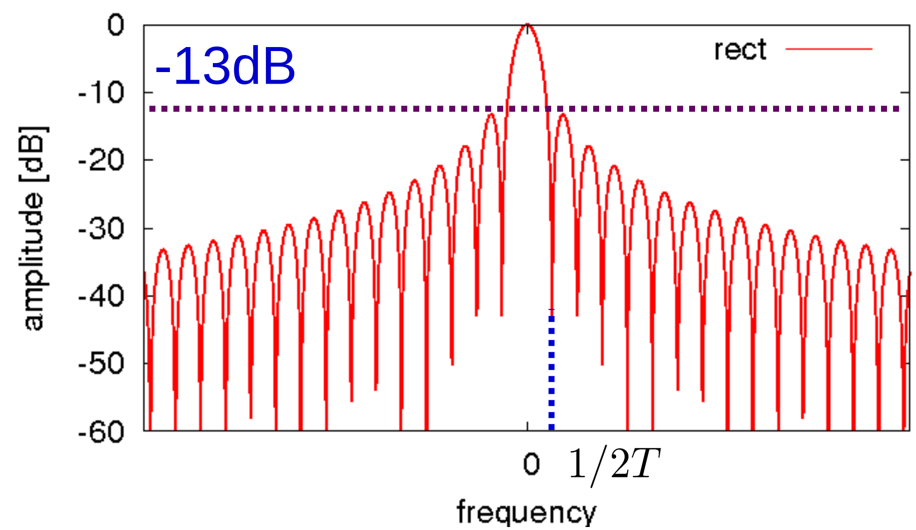


矩形窓

$$w(t) = \begin{cases} 1 & (|t| < T) \\ 0 & (|t| > T) \end{cases}$$

- 方形窓、Dirichlet窓とも呼ばれる
- メインローブの幅は窓関数中最も狭い
- サイドローブの最大値は-13dBと大きい

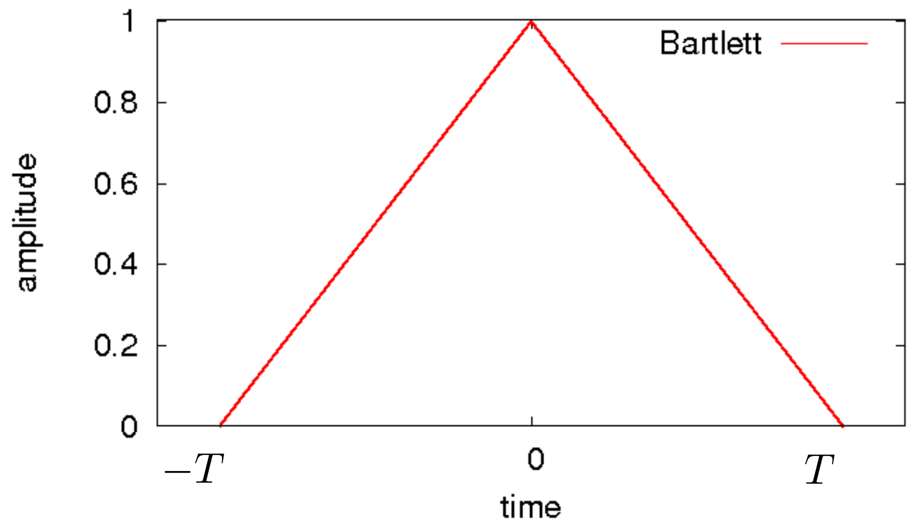


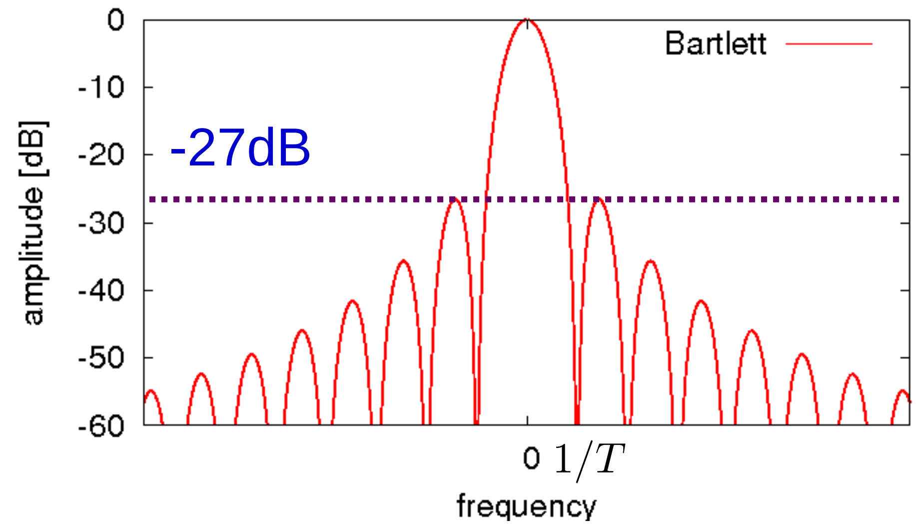


三角窓

$$w(t) = \begin{cases} 1 - \frac{|t|}{T} & (|t| < T) \\ 0 & (|t| > T) \end{cases}$$

- Bartlett窓とも呼ばれる
- 矩形窓の自己相関で表される
- サイドローブの最大値：-27dB

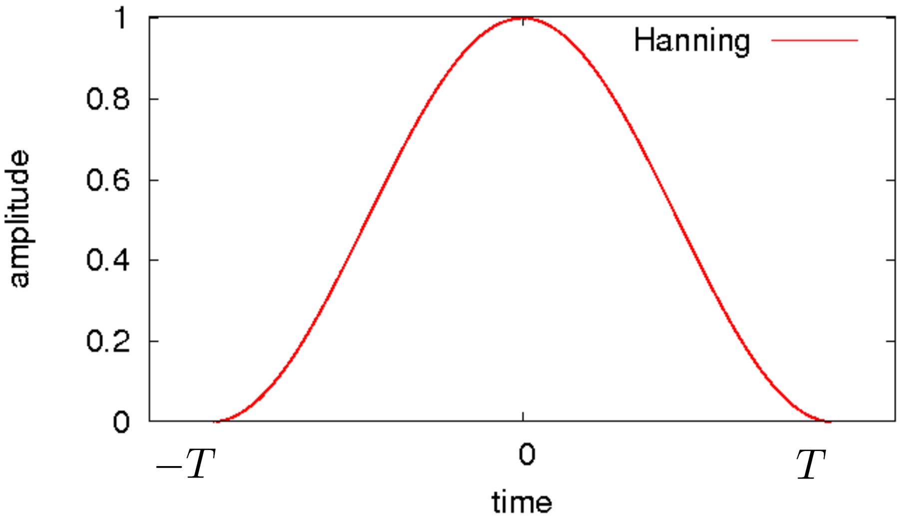


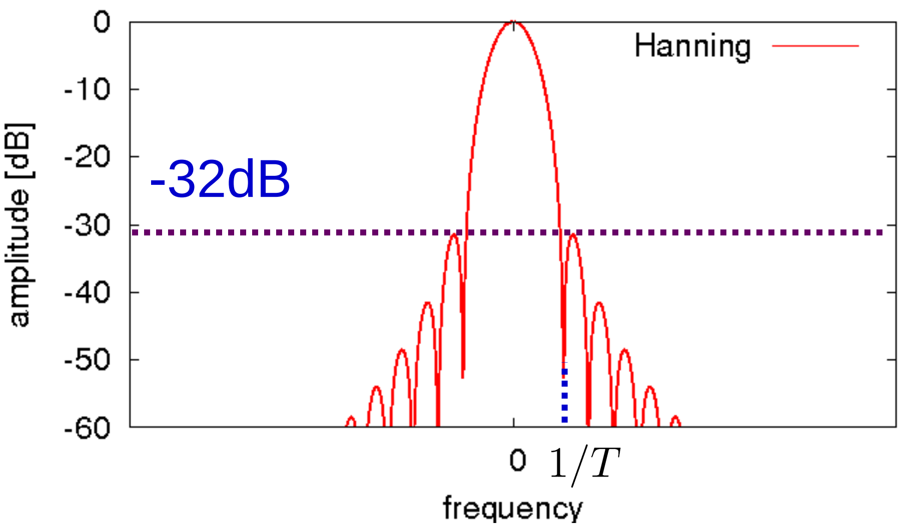


Hanning窓

$$
w(t) = \begin{cases} 0.5 + 0.5\cos\left(\dfrac{\pi t}{T}\right) & (|t| < T) \\ 0 & (|t| > T) \end{cases}
$$

- J. von. Hanにより提案
- サイドローブの最大値：-32dB
- メインローブの幅は三角窓より大きい

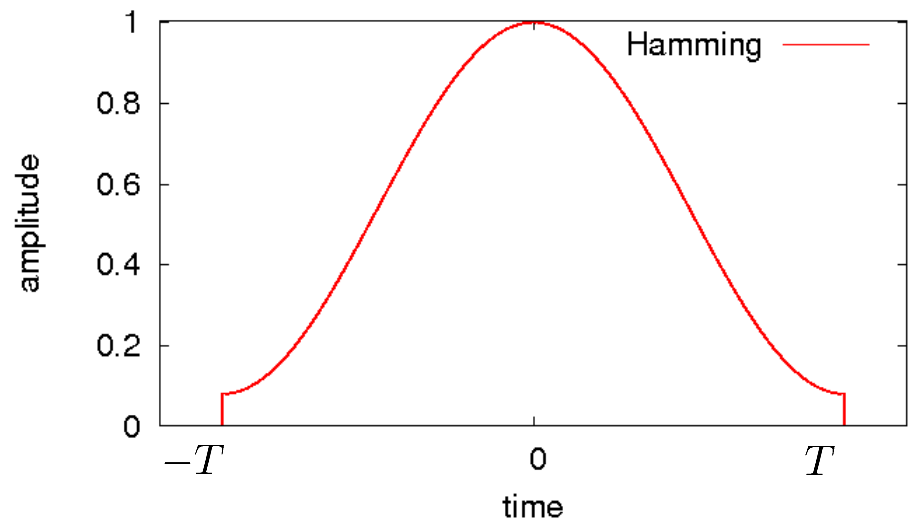


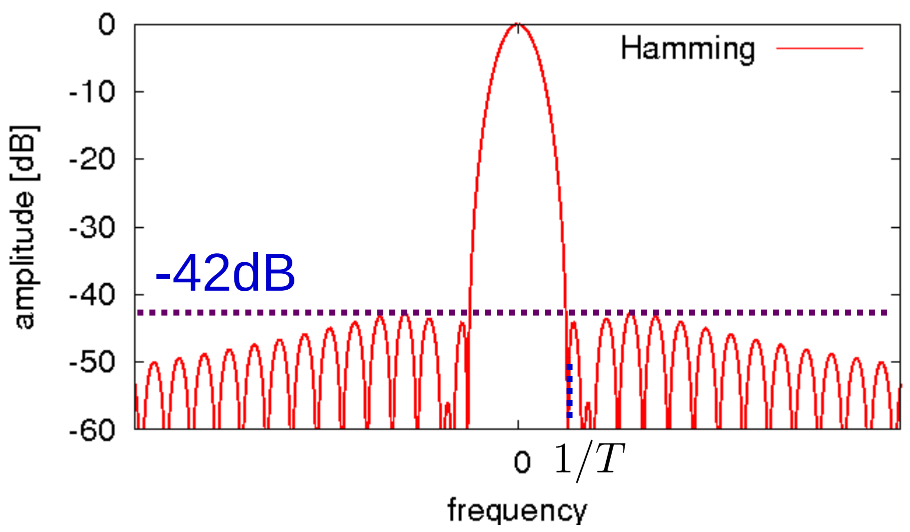


Hamming窓

$$w(t) = \begin{cases} 0.54 + 0.46\cos\left(\dfrac{\pi t}{T}\right) & (|t| < T) \\ 0 & (|t| > T) \end{cases}$$

- R. W. Hammingにより提案
- 両端で0にならない窓関数
- サイドローブの最大値：-42dB
- ただしサイドローブの減衰はHanning窓に比べ緩やか

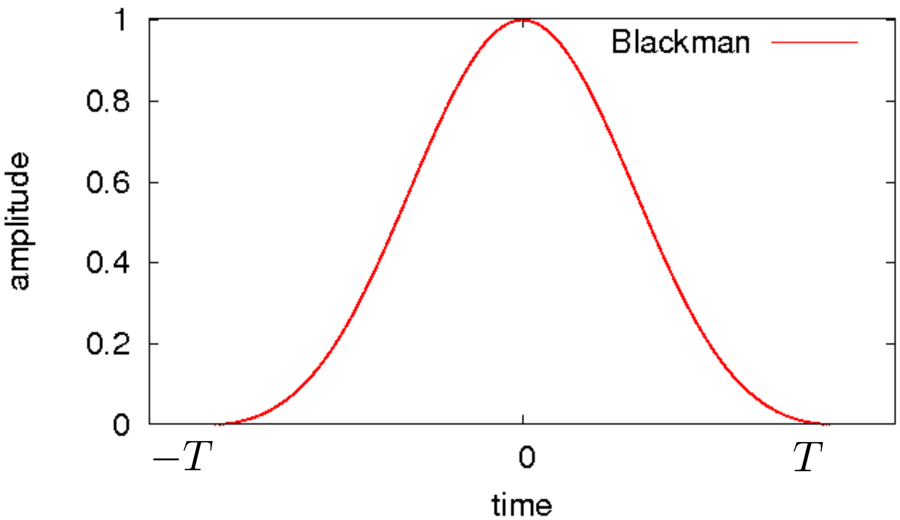


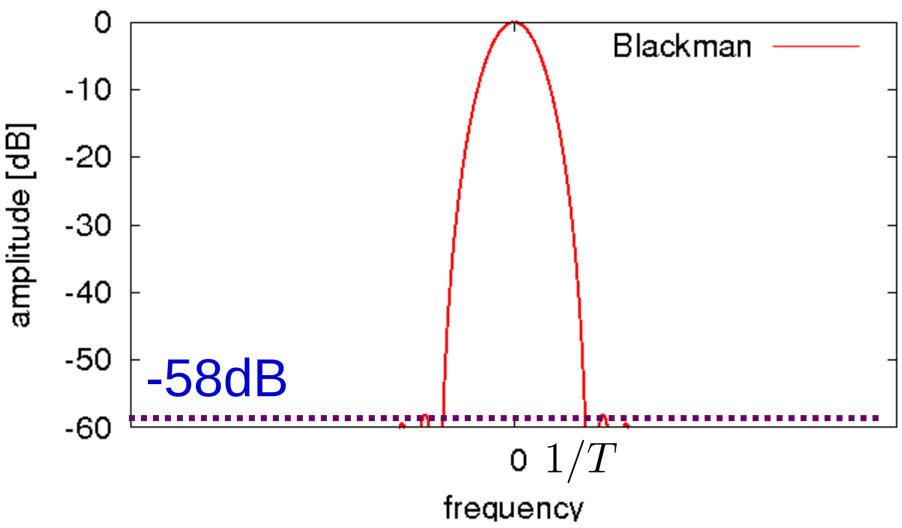


Blackman窓

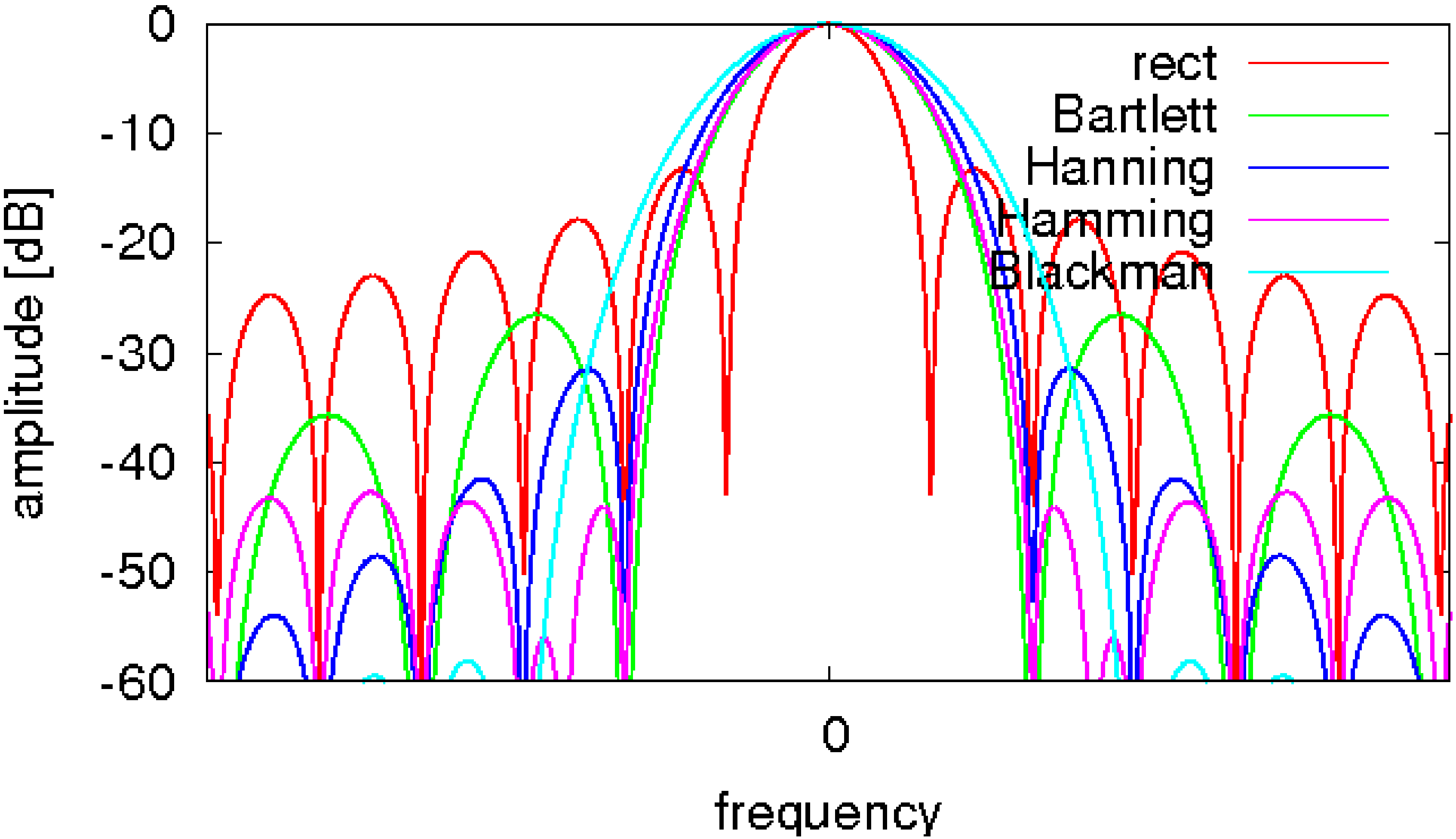
$$w(t) = \begin{cases} 0.42 + 0.5\cos\left(\dfrac{\pi t}{T}\right) \\ \qquad + 0.08\cos\left(\dfrac{2\pi t}{T}\right) & (|t| < T) \\ 0 & (|t| > T) \end{cases}$$

- R. Blackmanにより提案
- サイドローブの最大値：-58dB
- ただしメインローブの幅は比較的大きい





窓関数同士の比較



0
-10
-20
-30
-40
-50
-60
amplitude [dB]
rect
Bartlett
Hanning
Hamming
Blackman
0
frequency

時間周波数解析（短時間スペクトル分析）

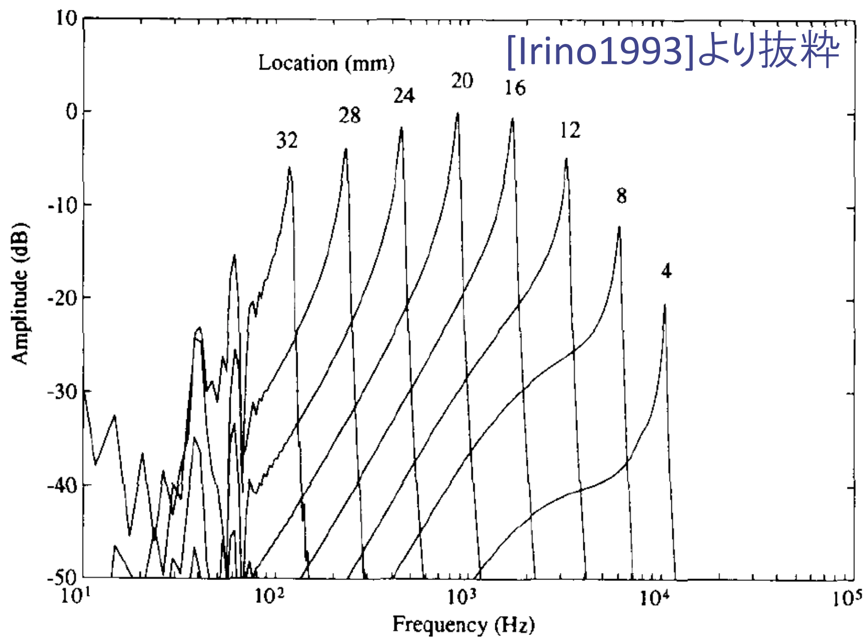
- 動機について
- 短時間Fourier変換 (Short Time Fourier Transform)
  - 定義
  - スペクトログラムとは
  - フィルタバンクとしての見方
- 聴覚フィルタバンク
  - 聴覚システムにおける時間周波数解析
  - 蝸牛モデル
- ウェーブレット変換（定Qフィルタバンク）
  - 定義
  - フィルタバンクとしての見方

時間周波数解析（短時間スペクトル分析）

- 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

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



聴覚伝達メカニズム（ムービー）

時間周波数解析（短時間スペクトル分析）

- 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

時間周波数解析(短時間スペクトル分析)

- 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

ウェーブレット変換（定Qフィルタバンク）

- 動機：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）

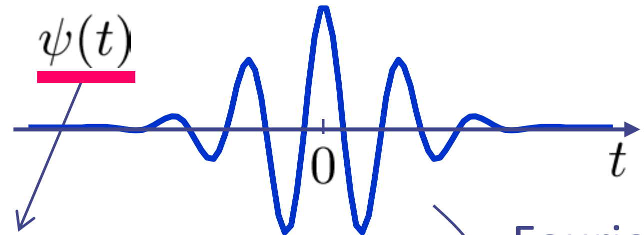
ウェーブレット変換（定Qフィルタバンク）

■<u>定義</u>：信号と「ウェーブレット」（小さい波）との内積

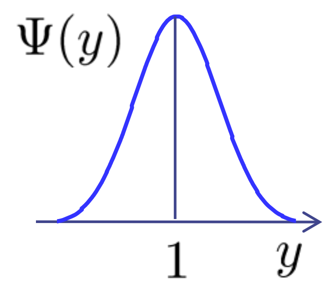
$$X_{\mathrm{wavelet}}(\alpha,\tau) = \langle x(t), \underline{\psi_{\alpha,\tau}(t)} \rangle_{t\in\mathbb{R}} = \int_{-\infty}^{\infty} x(t)\psi^*_{\alpha,\tau}(t)\mathrm{d}t$$

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

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）



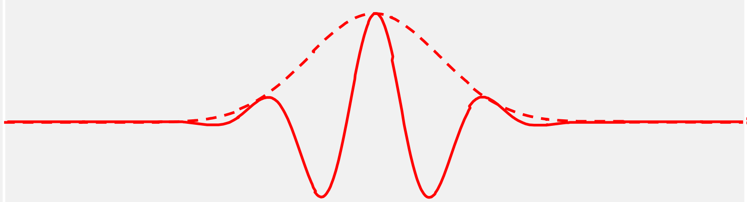
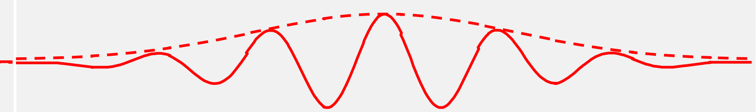
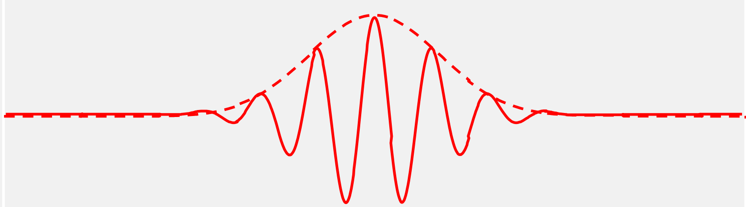
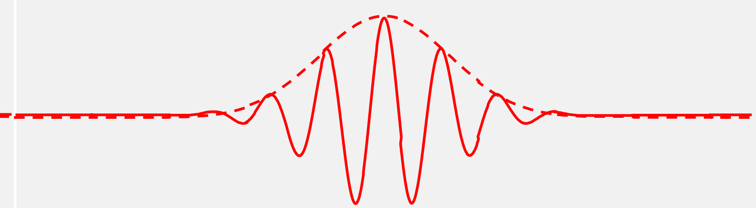
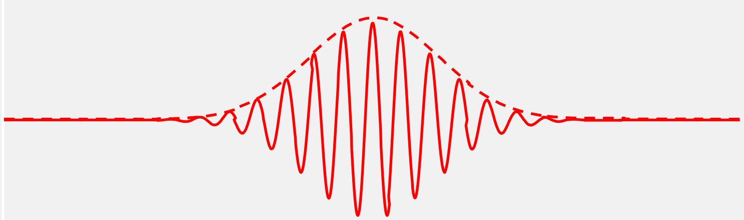
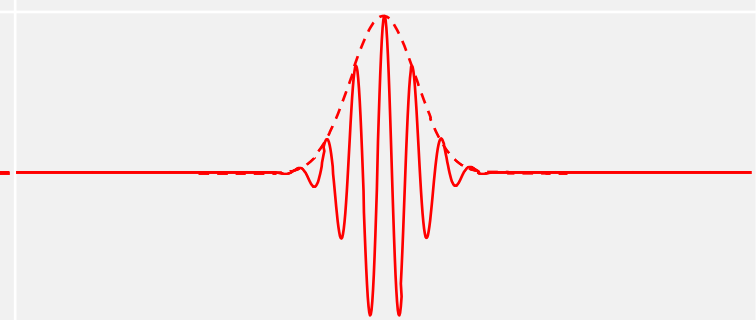
$\underline{\psi_{\alpha,\tau}}$ はアナライジングウェーブレットを時間方向に α 倍引き伸ばして，τ だけシフトしたもの

（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」？）

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

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

$\psi_{\alpha,\tau}$ 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

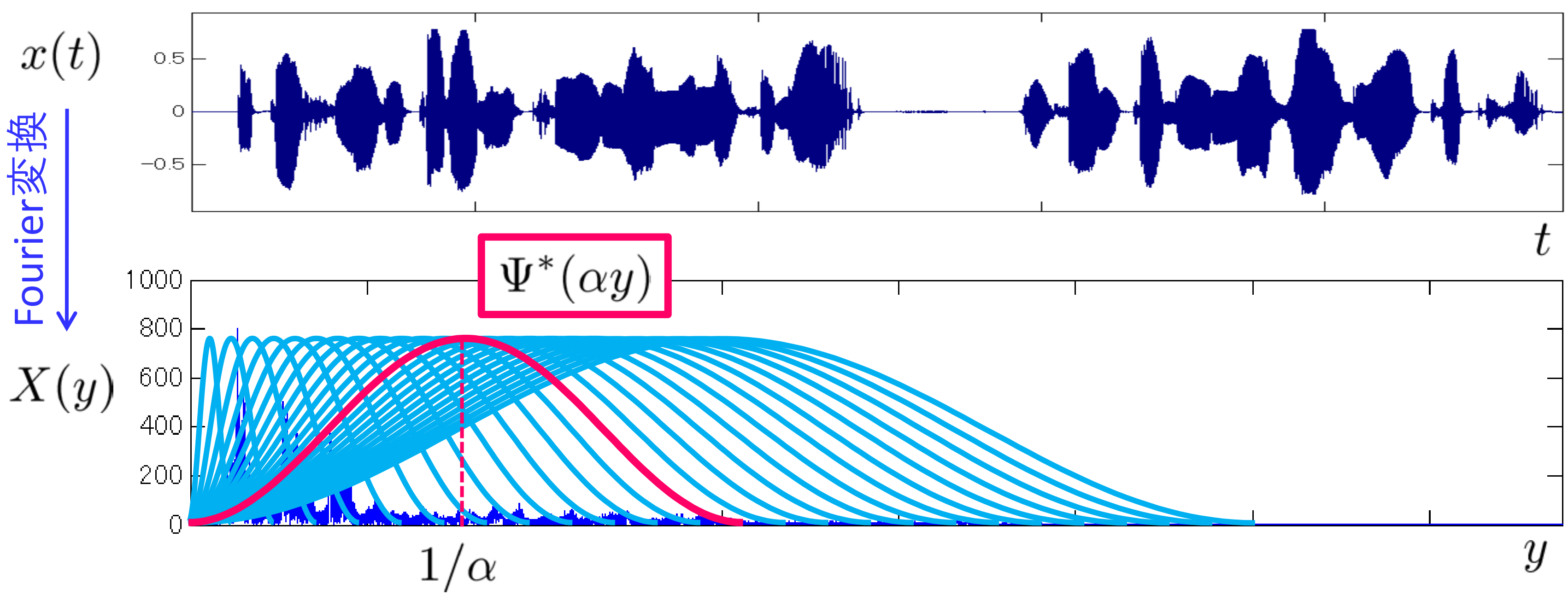
$$= \int_{-\infty}^{\infty} X(y)\Psi^*(\alpha y)e^{jy\tau}\mathrm{d}y$$

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

$\Psi_{\alpha,\tau}(y) = \underline{\Psi}(\alpha y)e^{-jy\tau}$

ψ のFourier変換

$X(y)\Psi^*(\alpha y)$ の逆Fourier変換



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