東京大学工学部 4年生 夏学期

応用音響学
第2回 (4/19)

猿渡　洋

東京大学大学院情報理工学系研究科
創造情報学/システム情報学専攻
hiroshi_saruwatari@ipc.i.u-tokyo.ac.jp

2019年度講義スケジュール

前半（猿渡担当）

- 4/05: 第1回
- 4/19: 第2回
- 4/26: 第3回
- 5/10: 第4回
- 5/17は休講予定
- 5/24: 第5回
- 5/31: 第6回

後半（小山先生担当）

- 6/07: 第7回
- 6/14: 第8回
- 6/21: 第9回
- 6/28: 第10回
- 7/05: 第11回
- 7/12は休講予定

- 7/26: 学期末試験（予定）

講義資料と成績評価

- 講義資料
  - http://www.sp.ipc.i.u-tokyo.ac.jp/~saruwatari/
    （システム情報第一研究室からたどれるようにしておきます）
- 成績評価
  - 出席点
  - 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

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

スパース最適化　ウィーナーフィルタ
ベイジアンフィルタ　非負値行列因子分解

非負値行列因子分解
隠れマルコフモデル

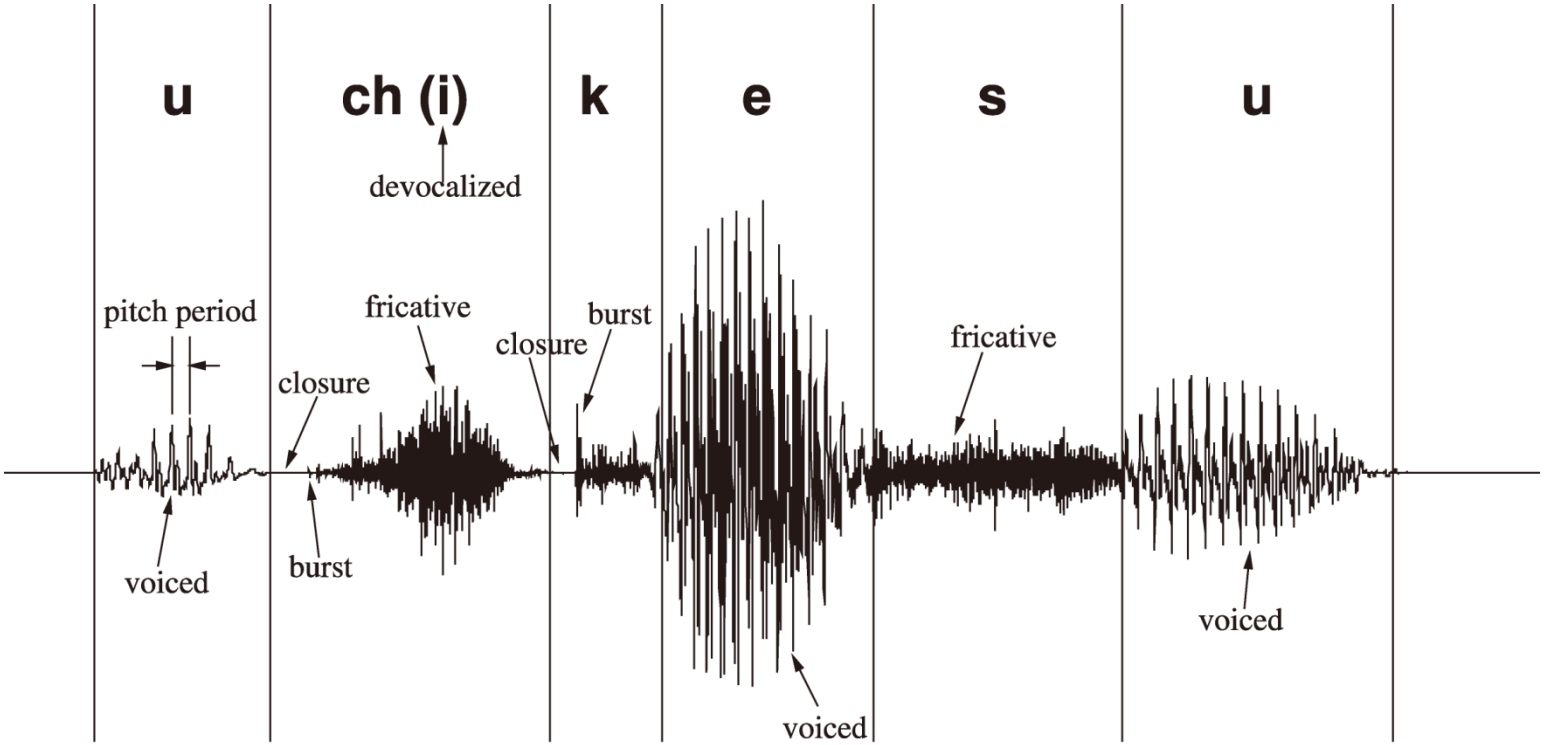
↑各分野の基礎トピック

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 (ShortTimeFourierTransform)
  - ウェーブレット変換（定Qフィルタバンク）
  - ケプストラム分析

音声波形

- 「打ち消す」の音声波形

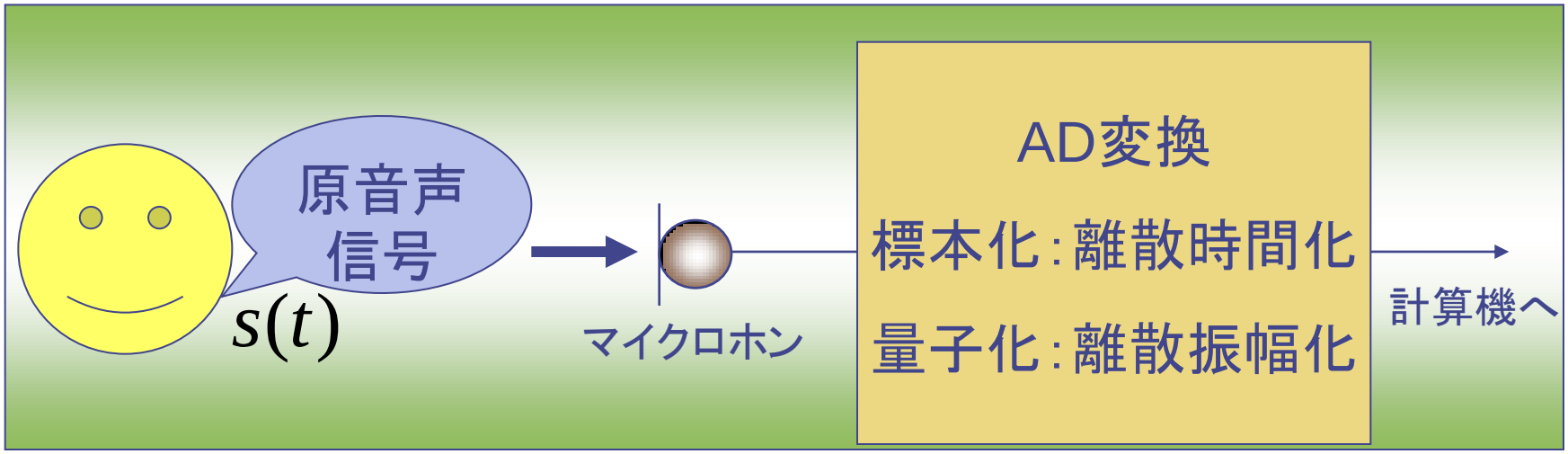


- 「ち」は無声化して母音の/i/が脱落
- 有声音（ここでは/u/, /e/, /u/）の波形は局所的に周期的
  →これがピッチ（声の高さ）として感じられる

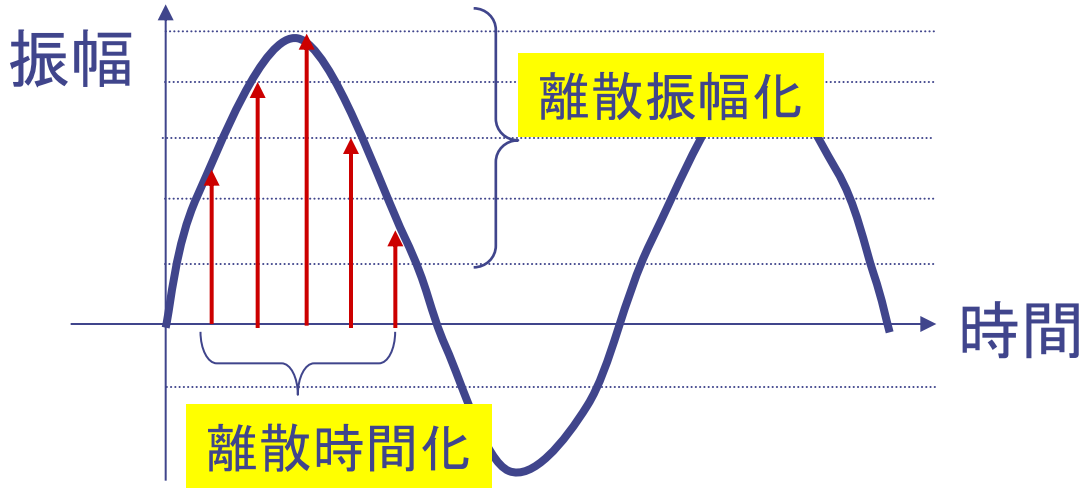
音声のディジタル化のプロセス

- 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

ディジタル音声処理の流れ

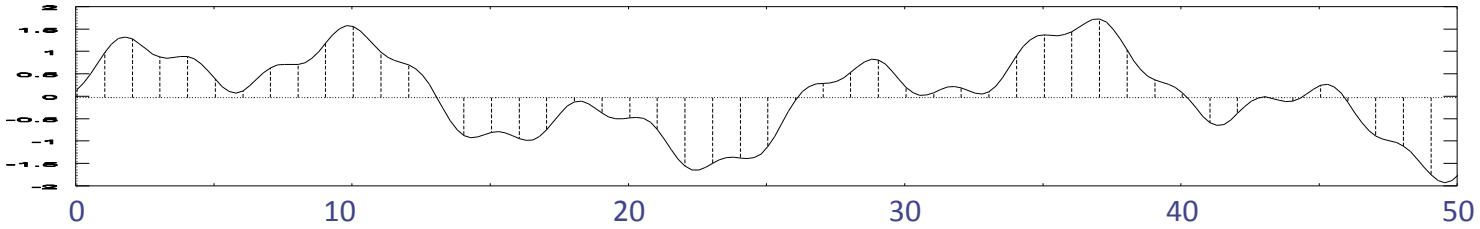


我々が実際に扱うことのできる信号は…

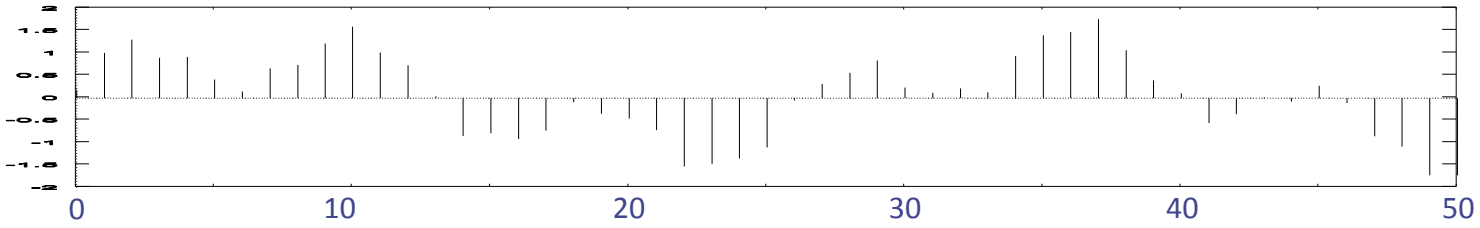


信号のサンプリング(標本化)と復元

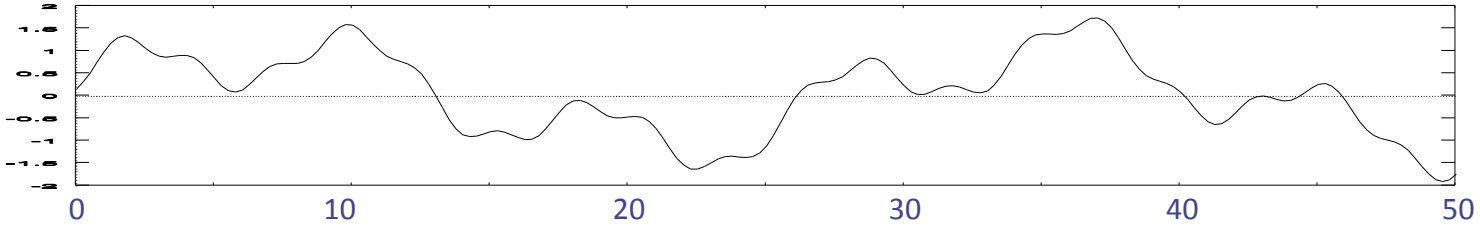
- 連続時間信号



- サンプリング(標本化)

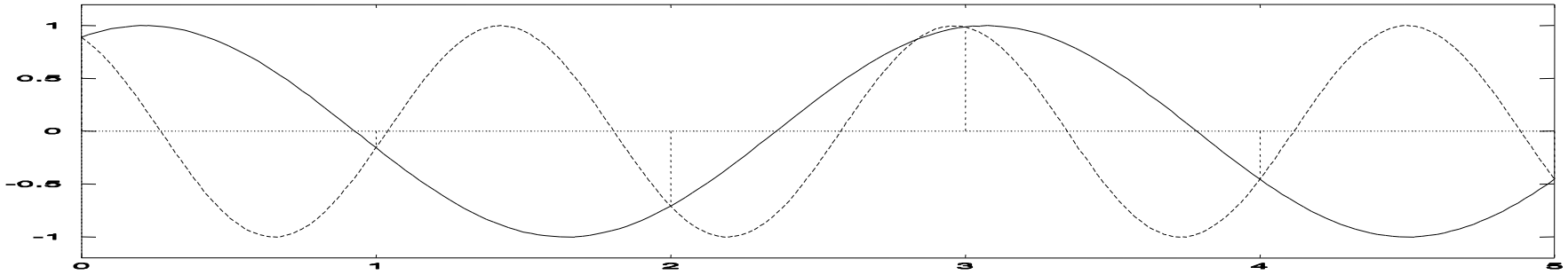


- 信号復元



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



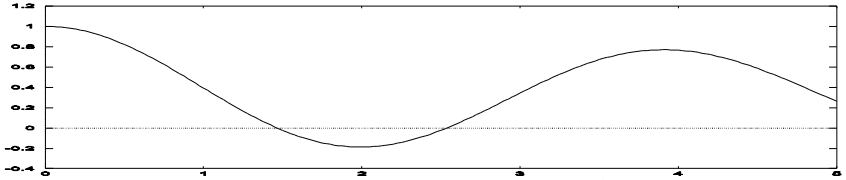
- 直感的には
  - 正弦波1周期あたり標本点が2点必要

標本化定理

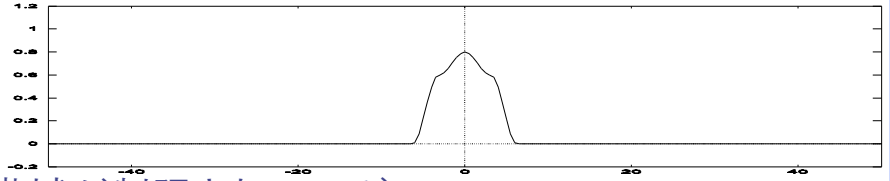
帯域制限信号に対し、カットオフ周波数の2倍以上で標本化すれば、標本値系列から元の信号が完全に復元される。

サンプリング定理(標本化定理)

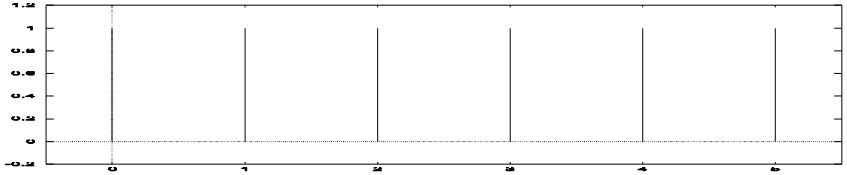
信号波形(時間領域)

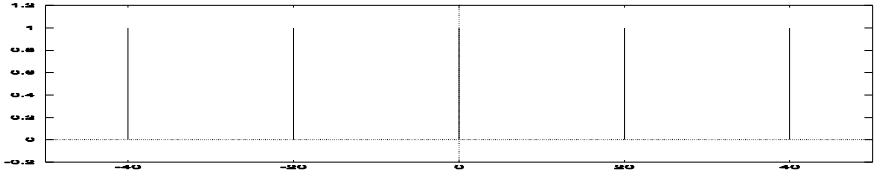


スペクトル(周波数領域)

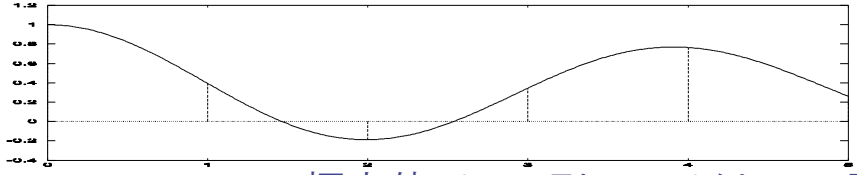


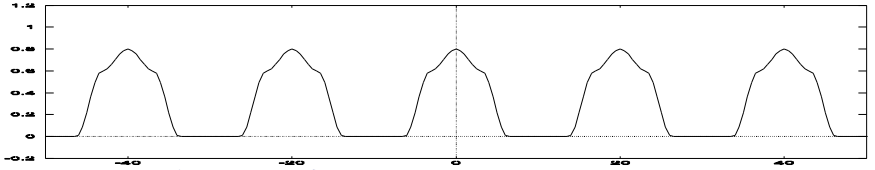
元の信号とスペクトル(帯域が制限されている)



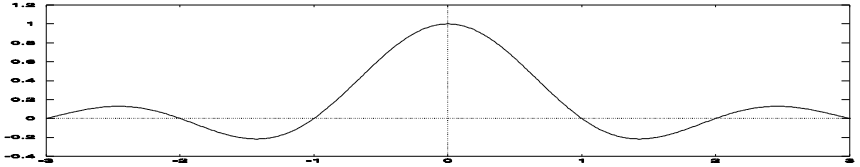


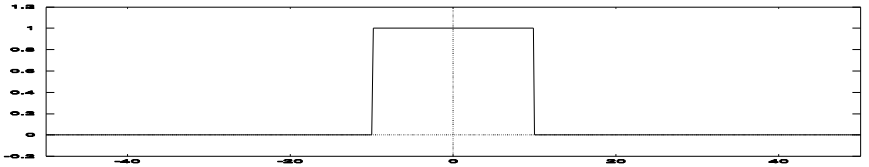
標本化＝元の信号にデルタ関数列を乗じること



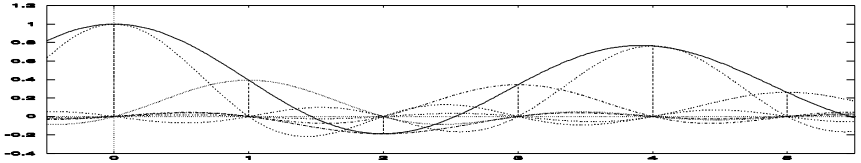


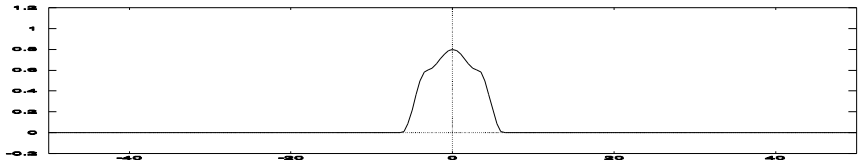
標本値パルス列のスペクトル＝原スペクトルとデルタ関数列の畳み込み





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

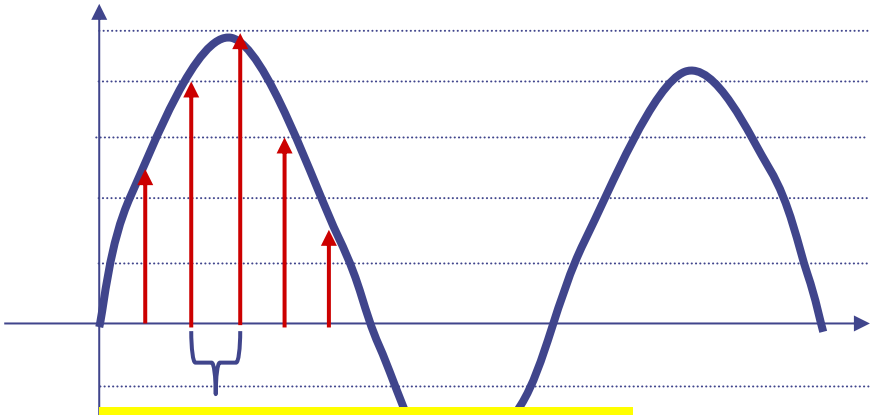




Shannon-染谷の標本化定理1

- アナログ信号波形$x(t)$が0～W [Hz]に帯域制限されているとき、$x(t)$をT＝1/(2W) [s]ごとに標本化すれば、標本値系列から以下のように波形再生を行うことができる。

$$x(t)=\sum_{i=-\infty}^{\infty} x\left(\frac{i}{2W}\right)\cdot\frac{\sin\{2\pi W(t-i/2W)\}}{2\pi W(t-i/2W)}$$



標本化周期T=1/(2W)

Shannon-染谷の標本化定理2

- $1/T = 2W$ [Hz]：　ナイキストレート

（例1）電話音声：　W =4 [kHz] ⇒T=1/8000 [s]

（例2）通常音声：　W =8 [kHz] ⇒T=1/16000 [s]

：

（例3）音楽信号：　W =20 [kHz] ⇒T=1/40000 [s]

（因みにCDは44100 [Hz]でサンプリングされている）

音声信号の量子化(1)

量子化ステップを Δ，量子化ビットを B，信号の振幅の存在範囲を L とすると，

$$\begin{aligned} \Delta 2^B &\geq L \\ B &\geq \log_2(\frac{L}{\Delta}) \end{aligned}$$

- 量子化によって生じる誤差 → 「雑音」とみなす

その量はどの程度と見積もればよいか？

1. 量子化雑音の平均パワーの算出：
   - e=量子化誤差，
   - $p(e)$=eの確率密度分布（$=1/\Delta$と近似）

$$\sigma_e^2 = \int_{-\Delta/2}^{\Delta/2} e^2 p(e) de = \int_{-\Delta/2}^{\Delta/2} e^2 (1/\Delta) de = \Delta^2/12$$

音声信号の量子化(2)

2. 信号の平均パワーとの比（SN比）の算出：

- σ_x=信号の実効値

→ 音声の場合は近似的に $L/2=4\sigma_x$

$$
\begin{aligned}
\text{SN比 [dB]} &= 10\log_{10}\frac{\sigma_x^2}{\sigma_e^2} = 10\log_{10}\frac{(L/8)^2}{\Delta^2/12} \\
&= 10\log_{10}\frac{\Delta^2\cdot 2^{2B}/64}{\Delta^2/12} \\
&\approx 6B - 7.2 \quad [\mathrm{dB}]
\end{aligned}
$$

1bit 量子化を細かくする → 約6 [dB] 雑音が減る

音声のダイナミックレンジは100 dBを超える→最低でも16 bitは必要

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
- ケプストラム分析
  - スペクトル包絡とピッチの分離

時間周波数解析の動機

- Fourier変換：

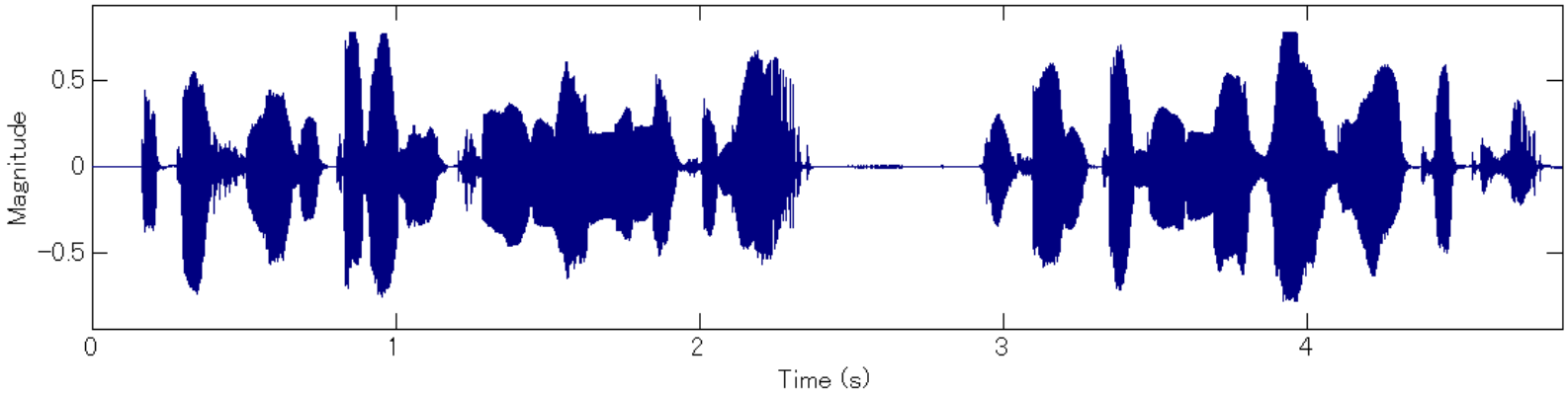
周波数ωの複素正弦波との内積

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

  - $|X(\omega)|$：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

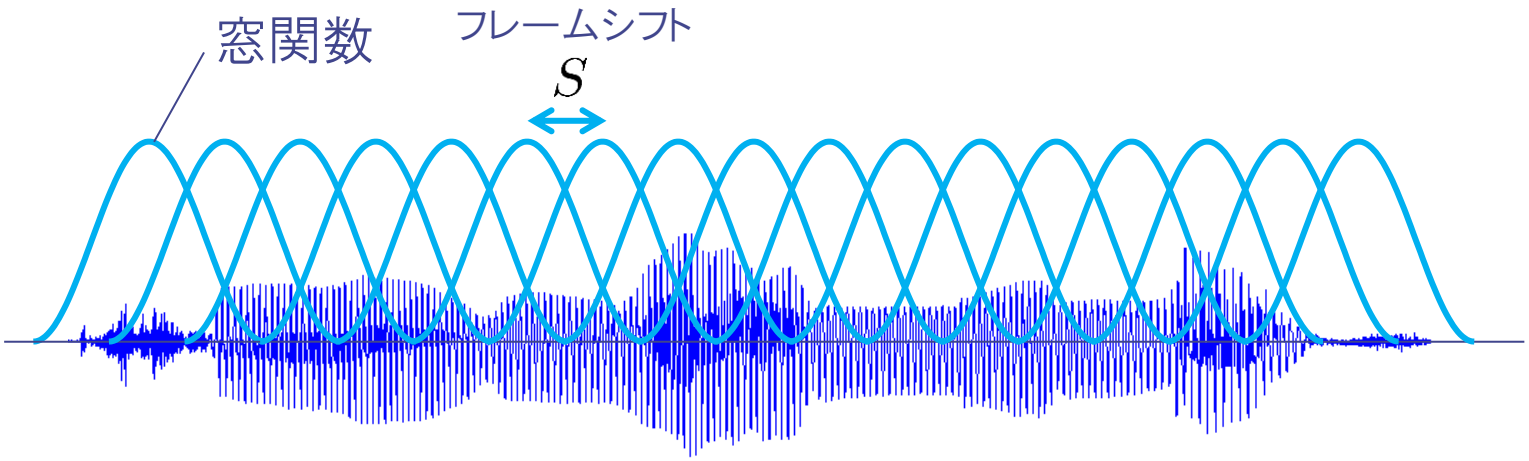
- 音声などの音響信号は非定常（非定常だから情報が運べる）



  - 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

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



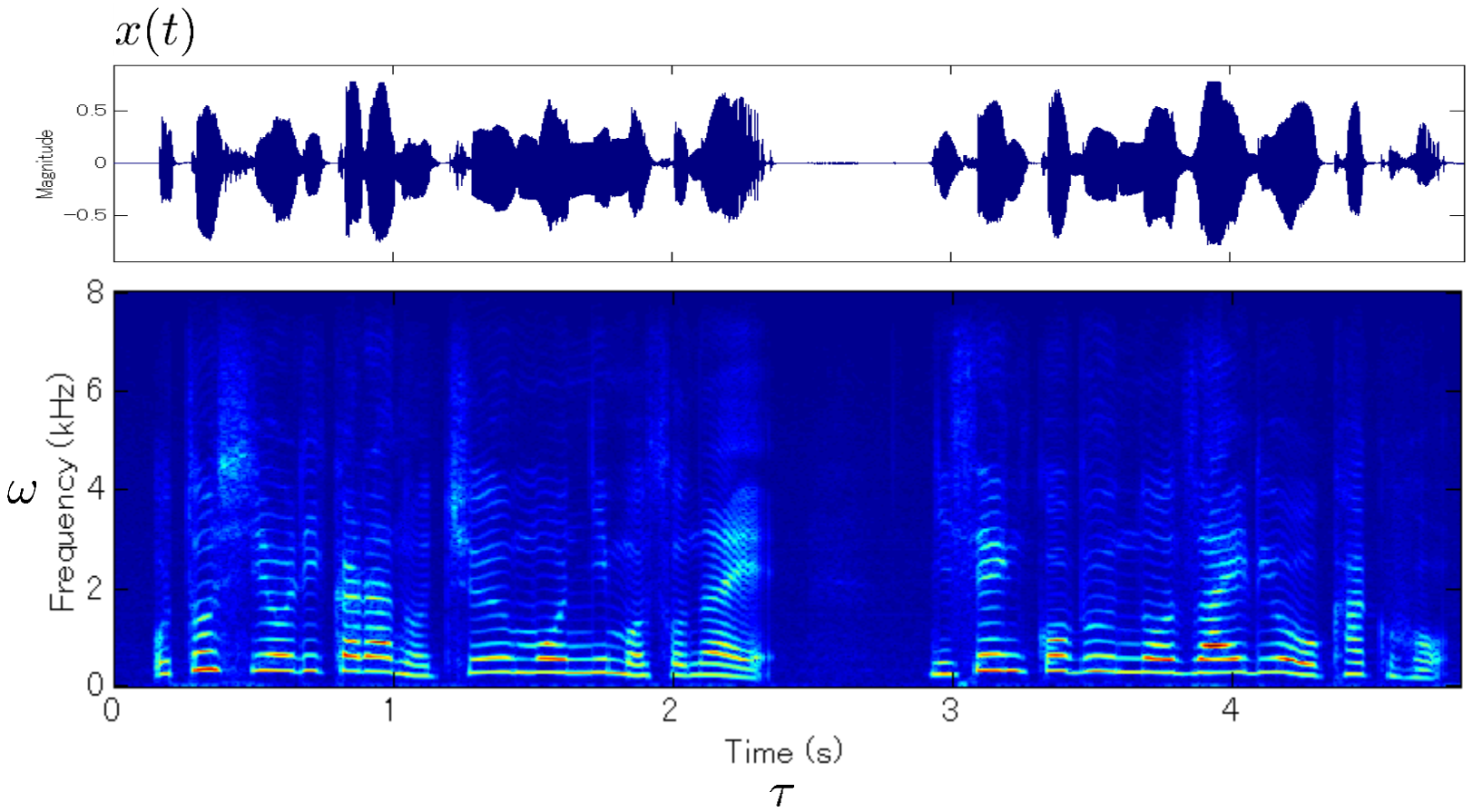
- 式で書くと・・・

m番目の窓で
切り出された波形

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

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)

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

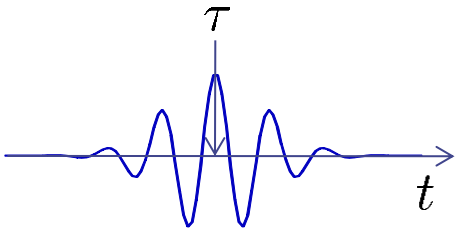
時刻 τ を中心とした窓で切り出された波形

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

時刻 τ に局在する周波数 ω の局在波

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

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



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

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

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

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

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

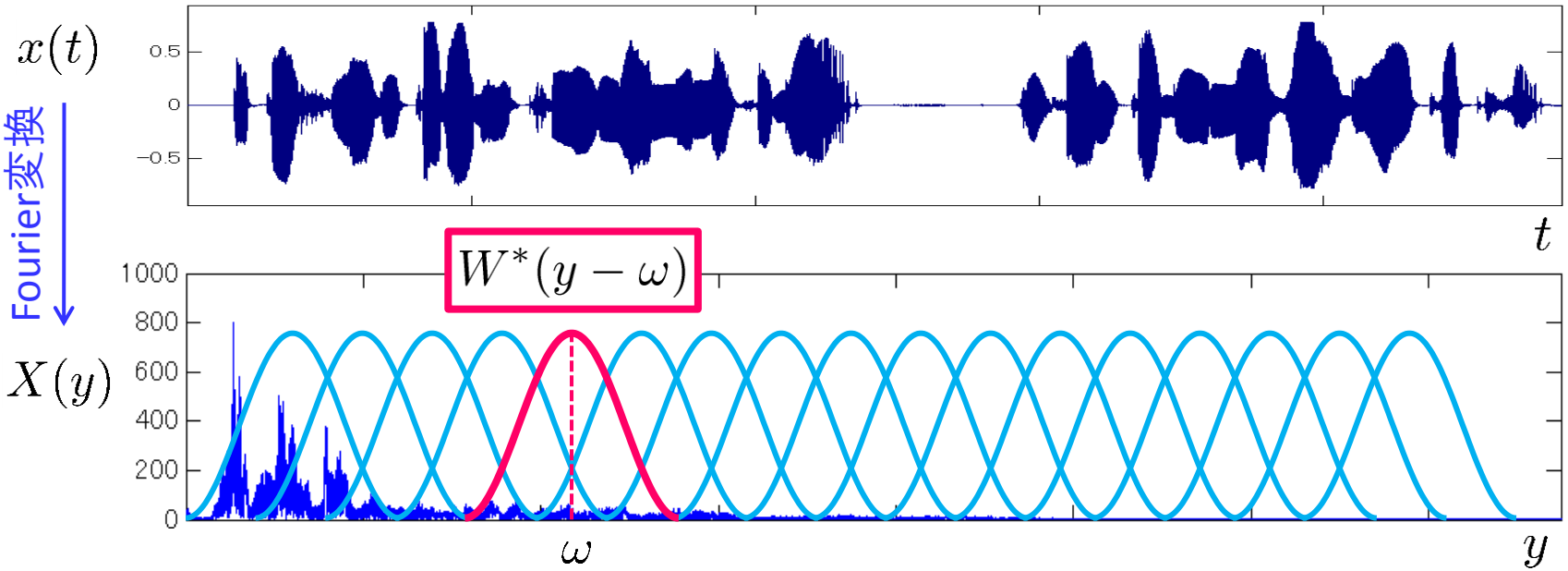
$$\Psi_{\omega,0}(y)=W(y-\omega)$$

w のFourier変換

フィルタバンクとしての見方 (2/2)

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

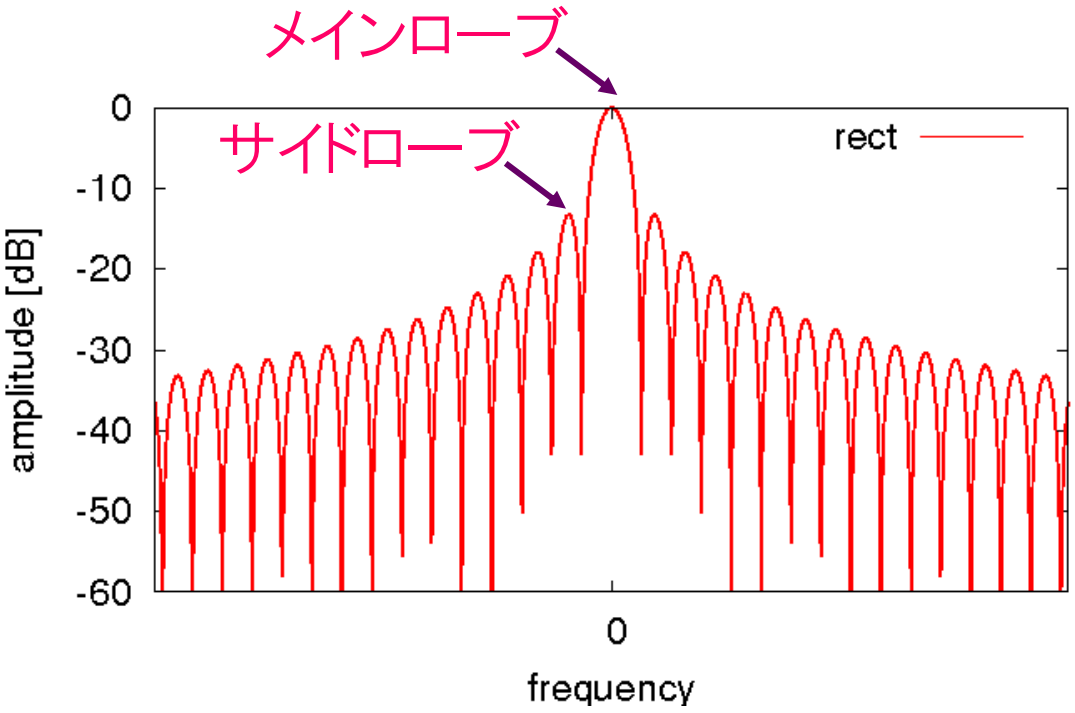
$X(y)W^{*}(y-\omega)$ の逆Fourier変換



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

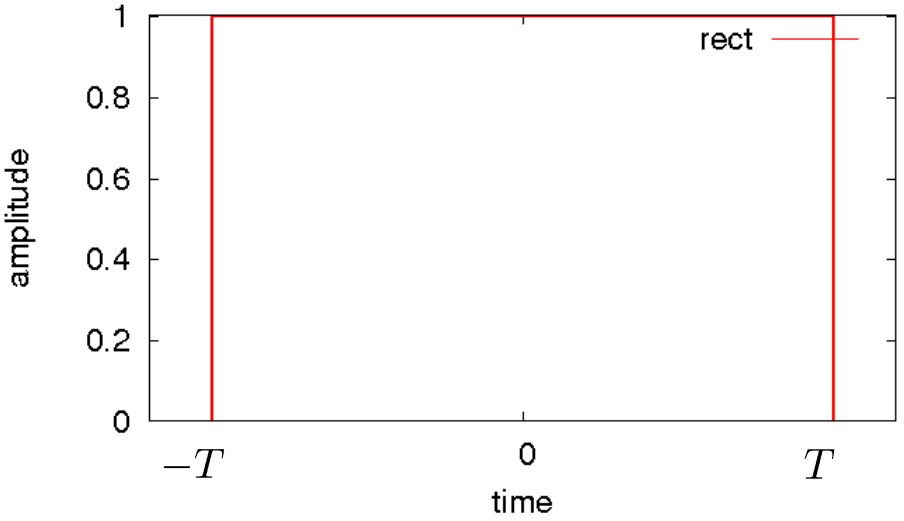
- 窓関数の重要な特性
  - メインローブの幅（周波数分解能）：狭いほどよい
  - サイドローブの大きさ（ダイナミックレンジ）：小さいほどよい
- 代表的な窓関数
  - 矩形窓
  - 三角窓（Bartlett窓）
  - Hanning窓
  - Hamming窓
  - Blackmann窓

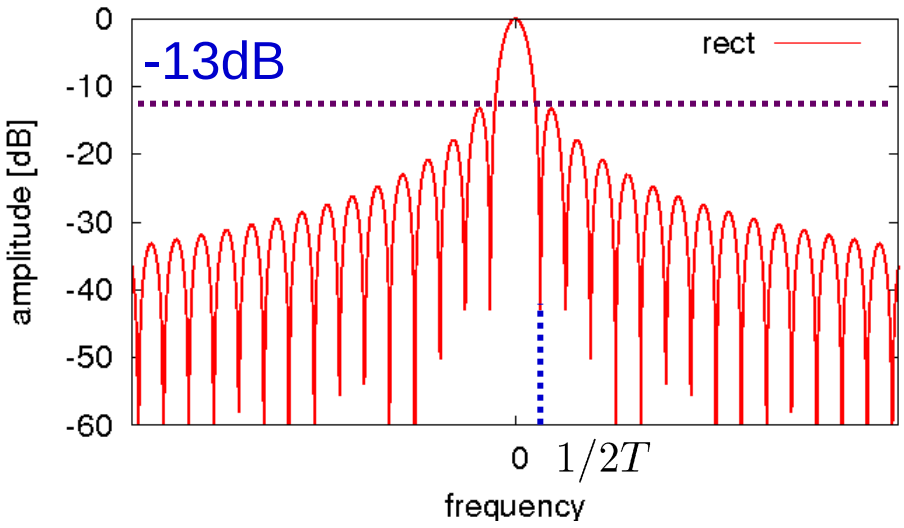


矩形窓

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

- 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

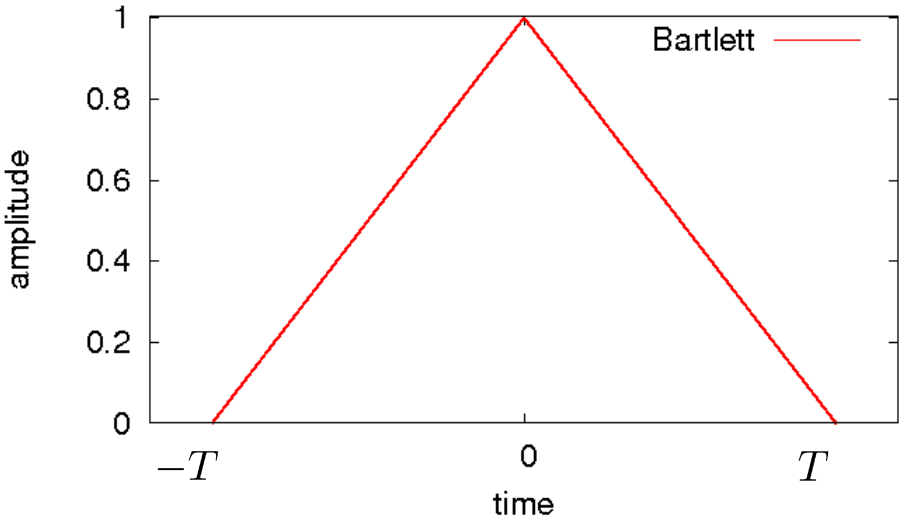


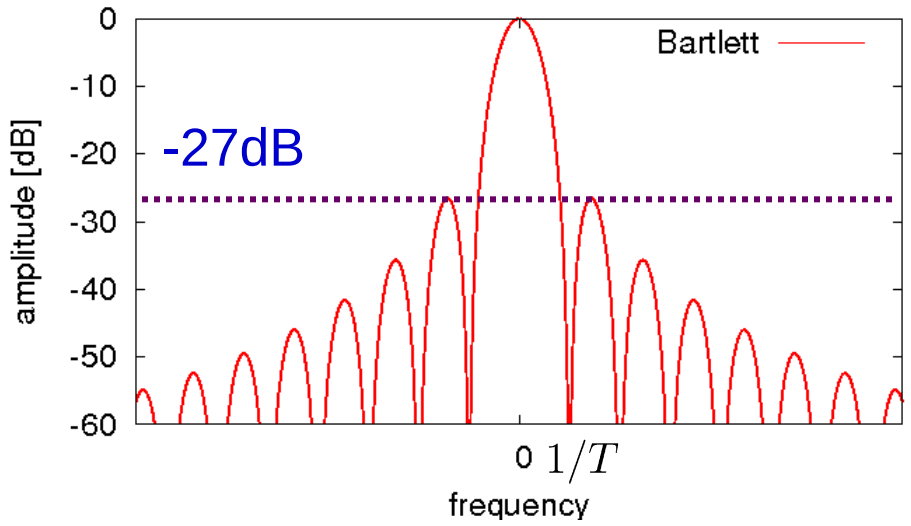


三角窓

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

- 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

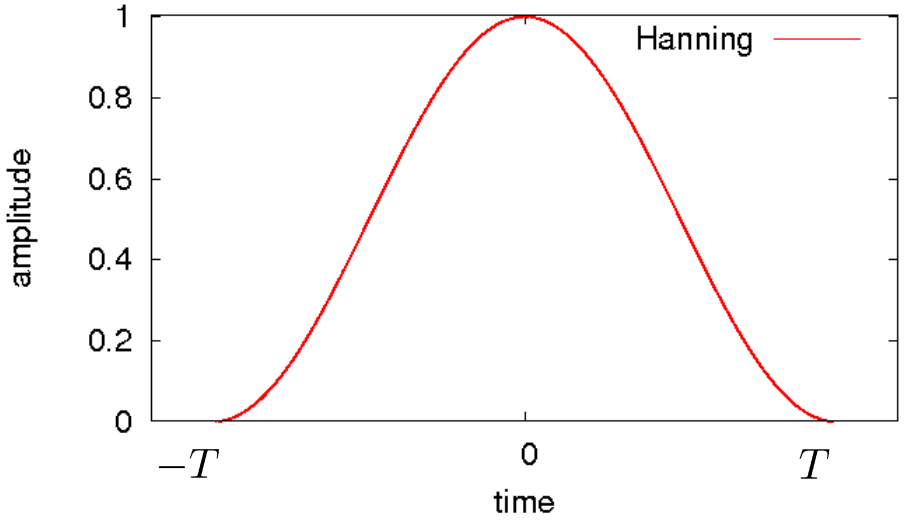


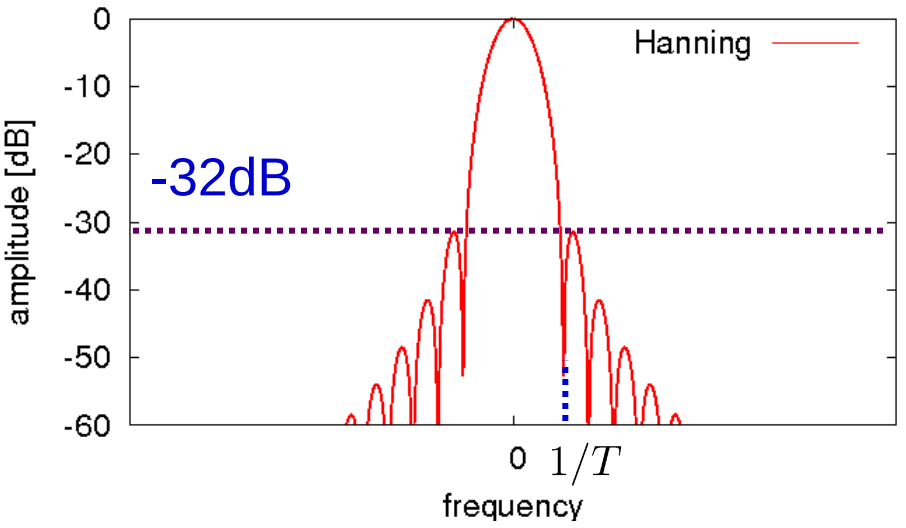


Hanning窓

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

- 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

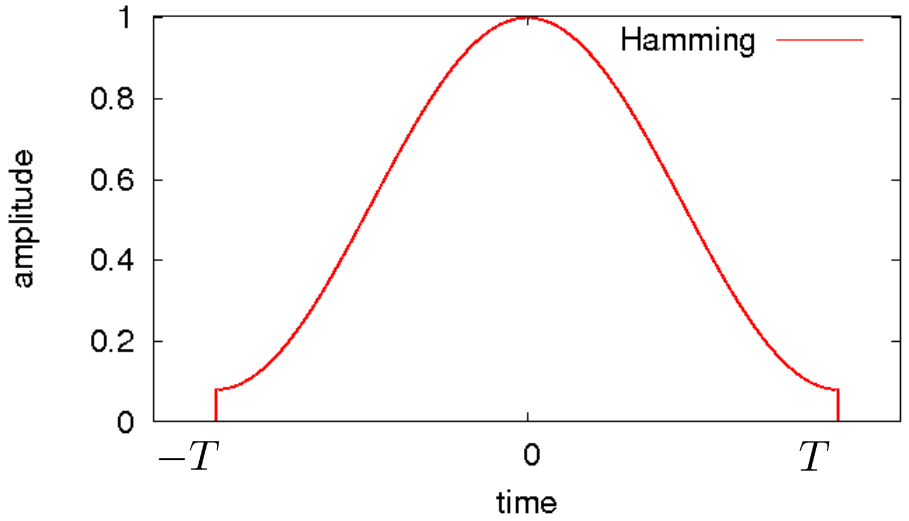


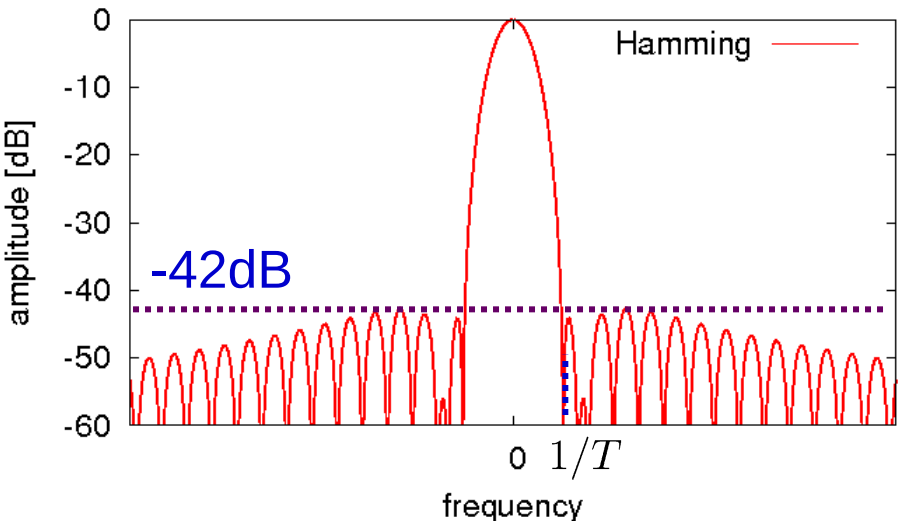


Hamming窓

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

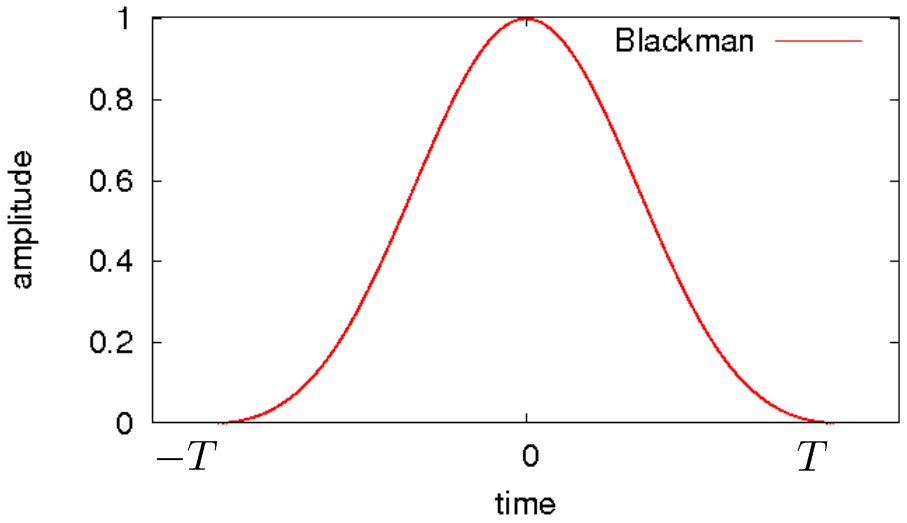
- R. W. Hammingにより提案
- 両端で0にならない窓関数
- サイドローブの最大値：
  -42dB
- ただしサイドローブの減衰は
  Hanning窓に比べ緩やか



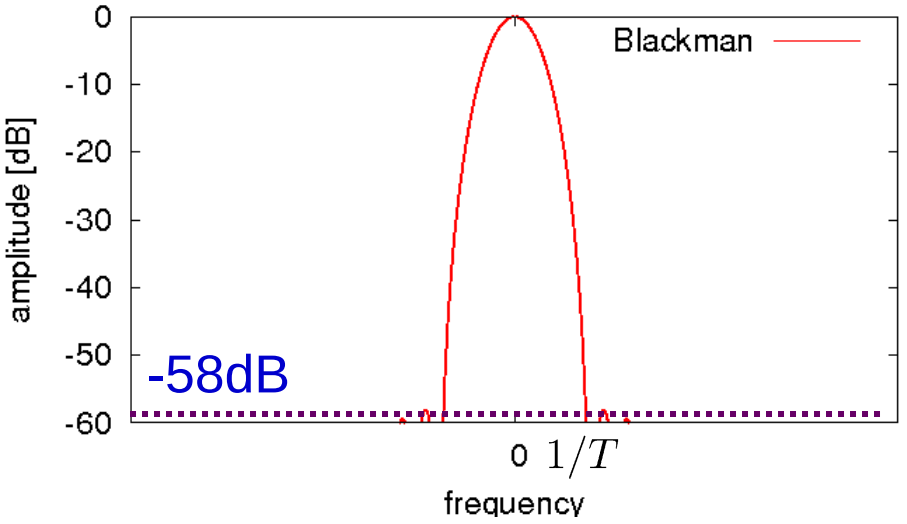


Blackman窓

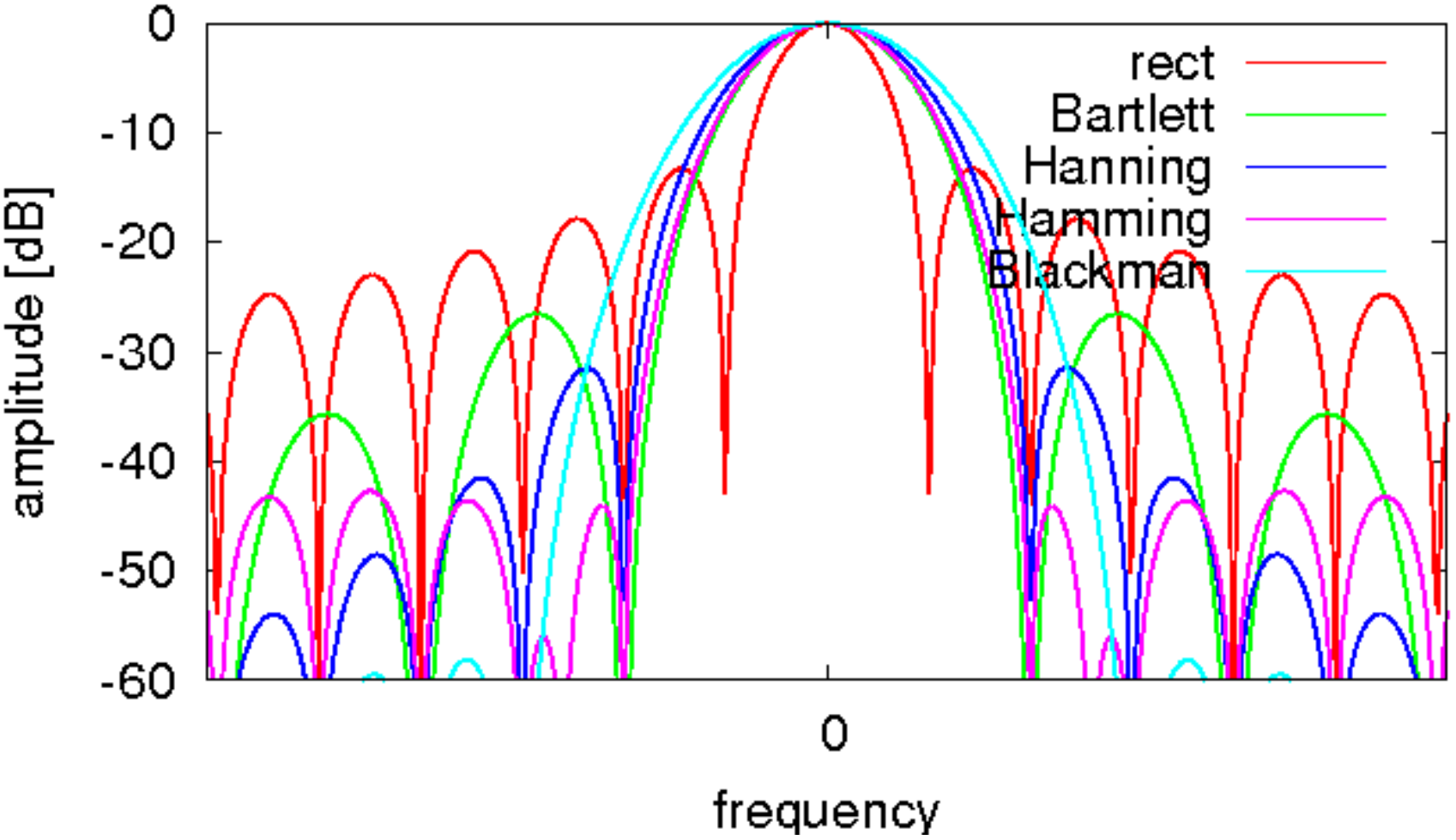
$$w(t) = \begin{cases} 0.42 + 0.5\cos\left(\frac{\pi t}{T}\right) \\ \qquad + 0.08\cos\left(\frac{2\pi t}{T}\right) & (|t| < T) \\ 0 & (|t| > T) \end{cases}$$



- 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]
0
frequency
rect
Bartlett
Hanning
Hamming
Blackman

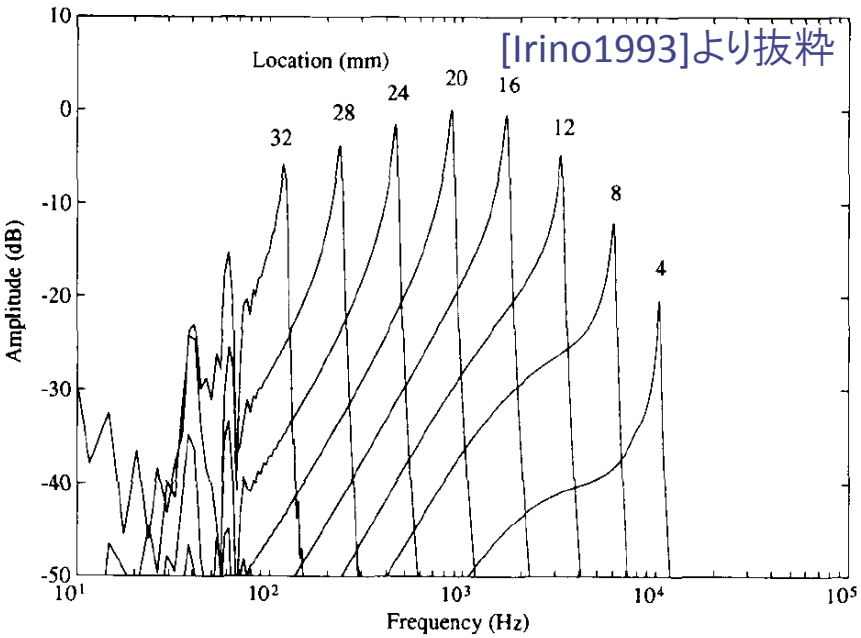
時間周波数解析（短時間スペクトル分析）

- 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
- ケプストラム分析
  - スペクトル包絡とピッチの分離

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



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）

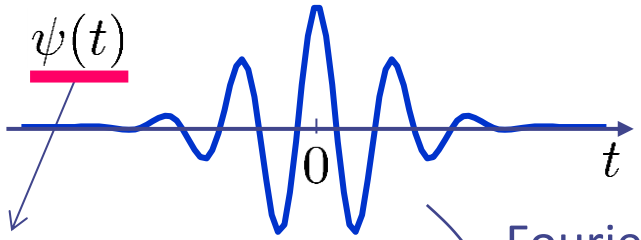
ウェーブレット変換（定Qフィルタバンク）

■定義：信号と「ウェーブレット」（小さい波）との内積

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

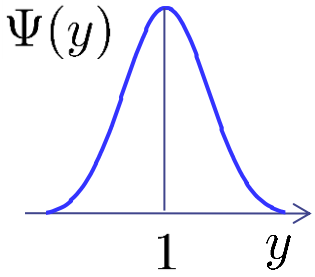
$$\psi_{\alpha,t}(t) := \frac{1}{\alpha}\psi\left(\frac{t-\tau}{\alpha}\right)$$

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



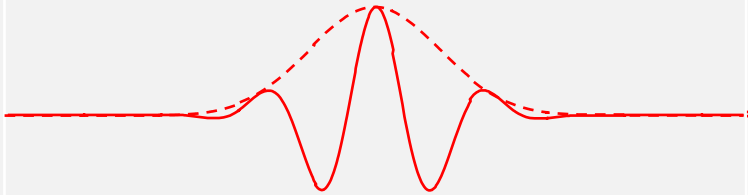
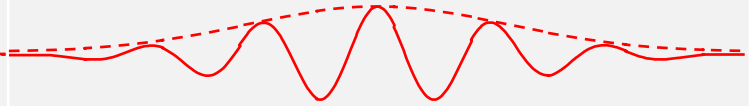
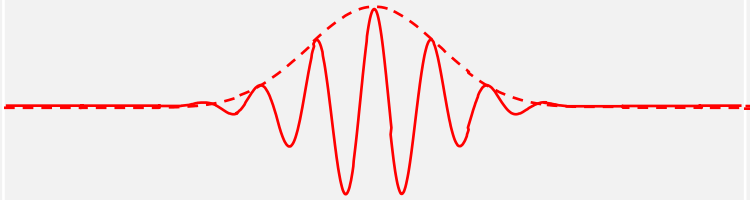
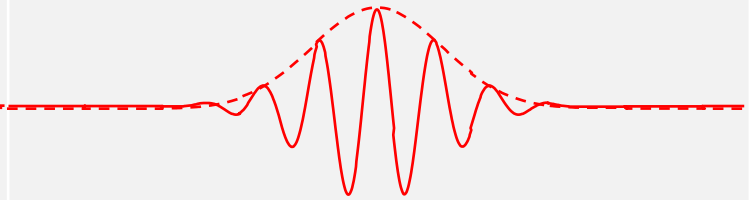
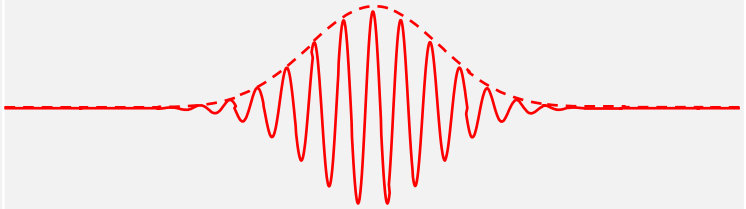
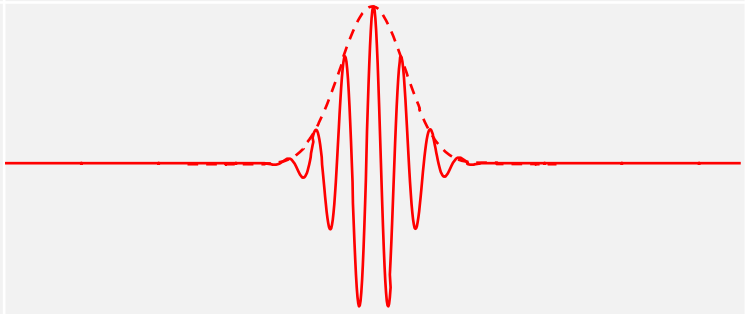
$\psi_{\alpha,\tau}$ はアナライジングウェーブレットを時間方向に α 倍引き伸ばして，τ だけシフトしたもの

（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」？）

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

$\psi_{\alpha,\tau}$ のFourier変換

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

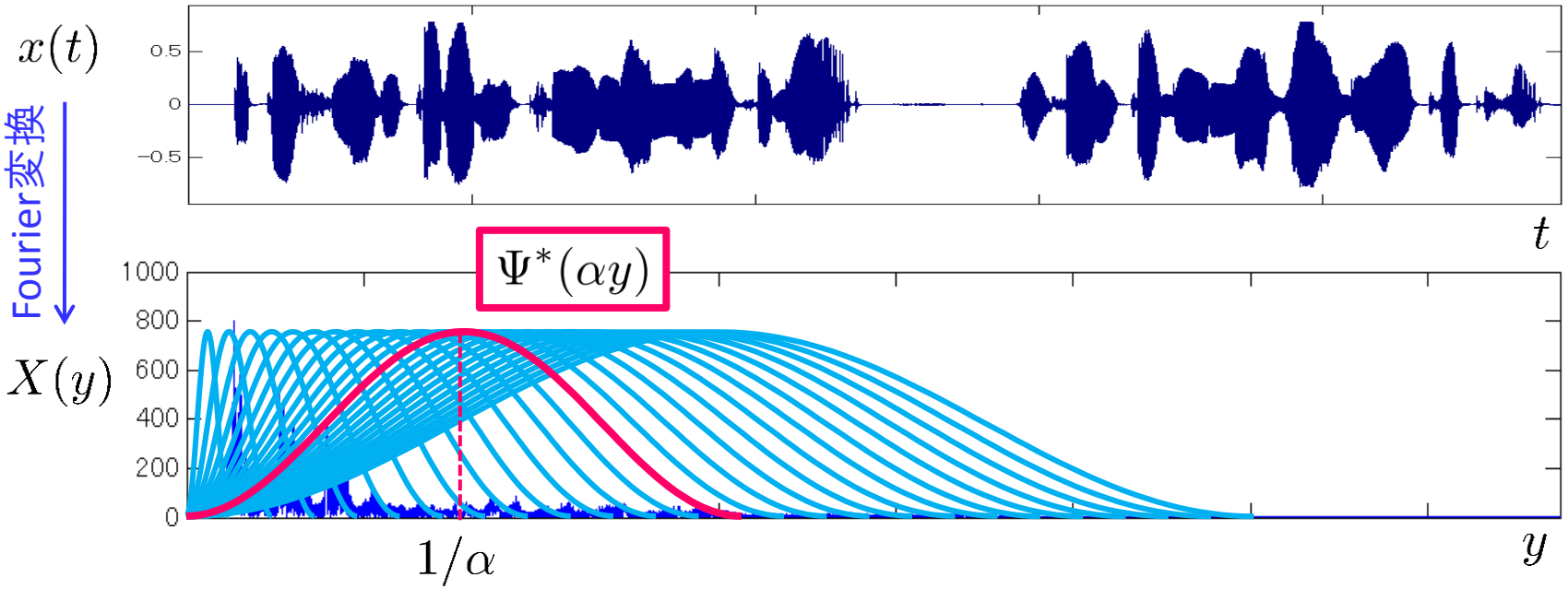
$X(y)\Psi^*(\alpha y)$ の逆Fourier変換

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

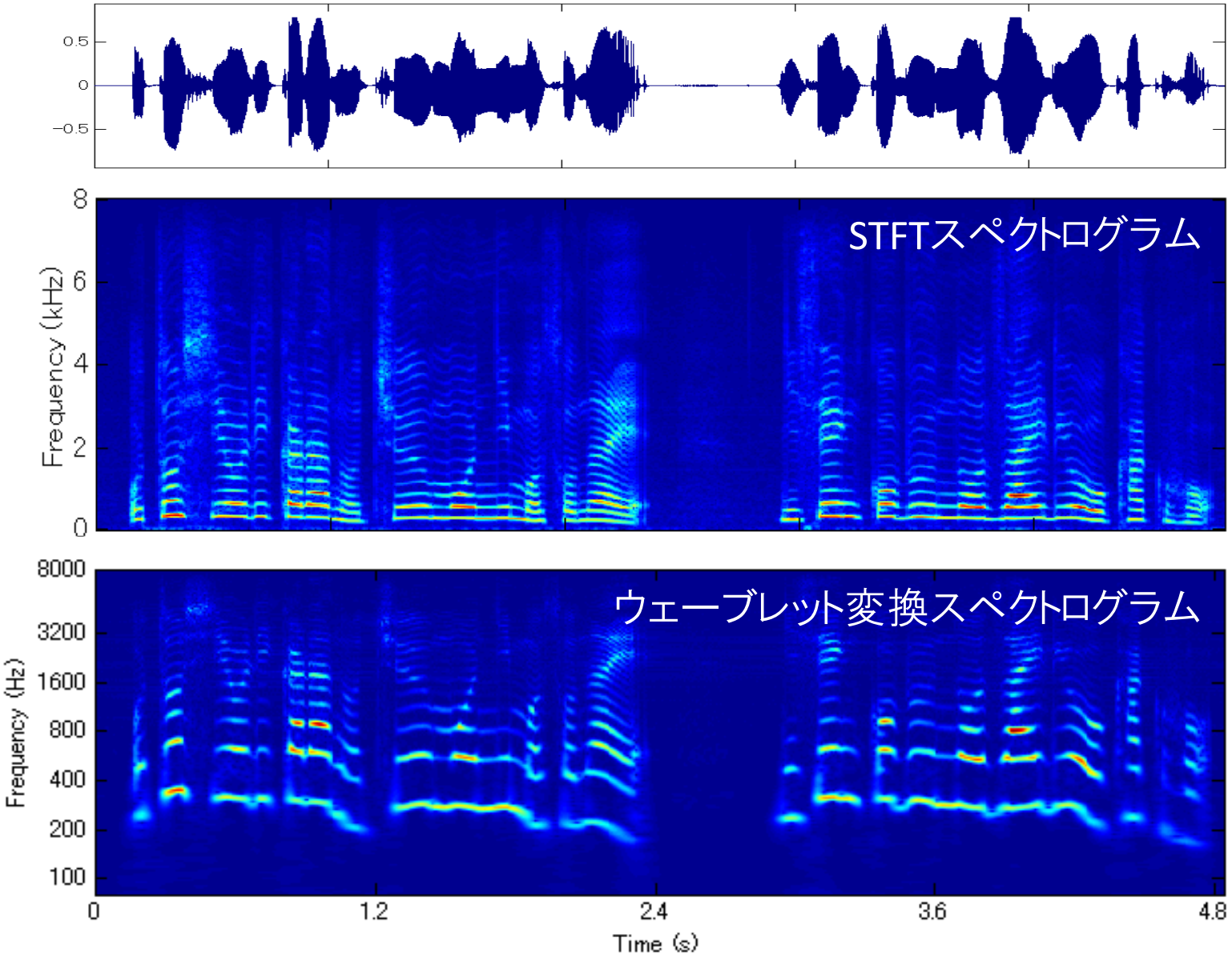
$\psi_{\alpha,t}(t) = \frac{1}{\alpha}\psi\left(\frac{t-\tau}{\alpha}\right)$ より

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

ψ の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



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
- ケプストラム分析
  - スペクトル包絡とピッチの分離

音声のスペクトル構造1

短時間スペクトル

- 音声は、短時間区間ごとの電力スペクトル密度（周波数領域におけるパワー特性）で測ることが多い。

音声スペクトル構造の2要素

- 周波数とともにゆるやかに変化する成分[スペクトル包絡]

  ⇒発声器官の共振・反共振特性を表す

  （つまり人間の喉・口の形をあらわす特徴量）

- 細かく周期的（有声音；母音などの場合）または非周期的（無声音の場合）に変化する成分　[スペクトル微細構造]

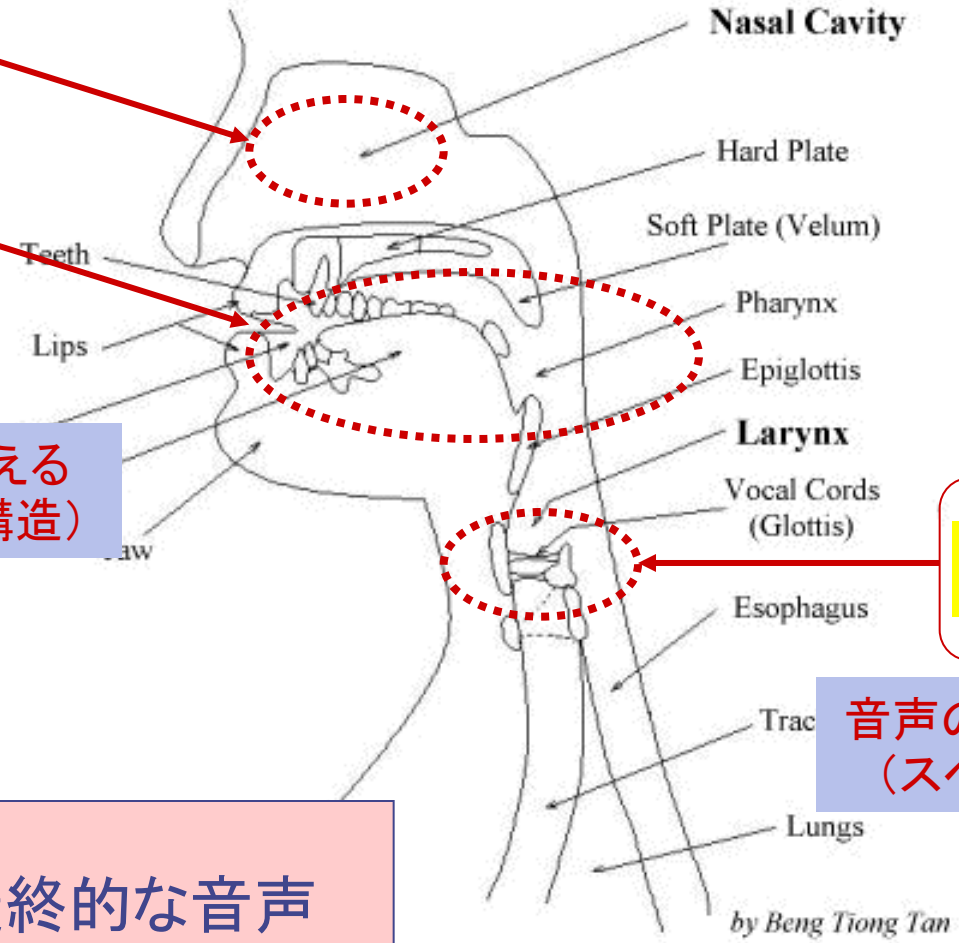
  ⇒音源の周期性

  （つまり声帯の基本周期・声の高低を表す特徴量）

音声信号のスペクトルはこれら2つの要素の積で表される

音声のスペクトル構造：発声構造

人間頭部の断面図



② 鼻腔 / 声道 の共振

音声の音色を与える
（スペクトル包絡構造）

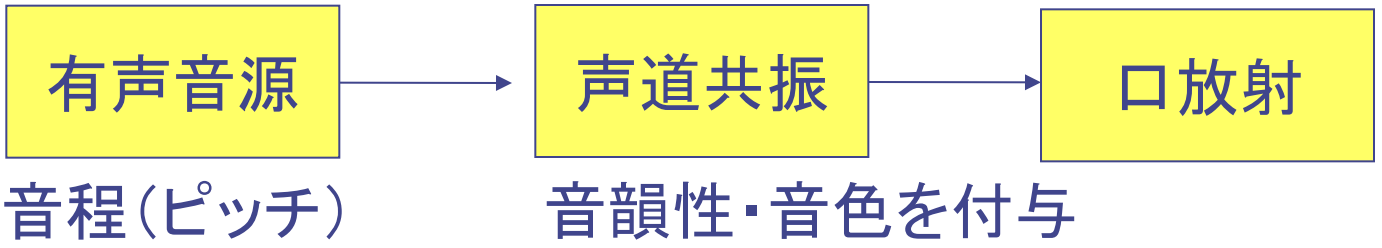
① 声帯 の振動

音声の基本周期を与える
（スペクトル微細構造）

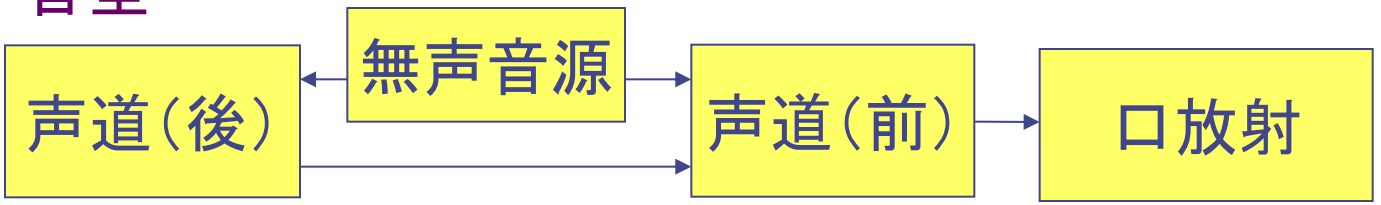
①×②＝最終的な音声

音声のスペクトル構造：発声構造2

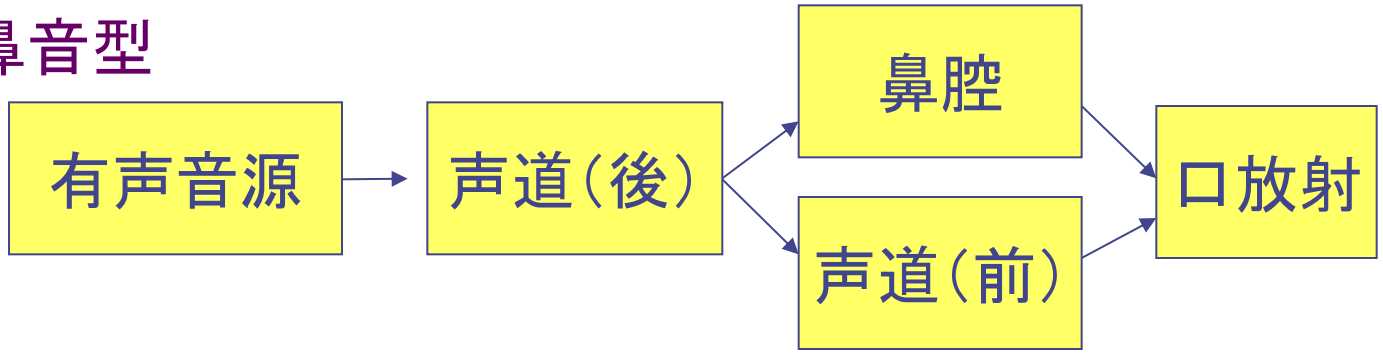
母音型



子音型



鼻音型



音声のスペクトル分析手法

- **短時間スペクトルを求める2手法**
  - ノンパラメトリック分析とパラメトリック分析

- ノンパラメトリック分析法：
  - 分析対象の信号に関して、特にモデルを仮定せずに周波数分析を行う手法。万能であるが抽出すべきパラメータは多くなる。
  - （例）短時間DFT（離散フーリエ変換）分析，ケプストラム分析
- パラメトリック分析法：
  - 分析対象信号に対して特定のモデル化を行い、そのモデルを表現する特徴パラメータを抽出する。音声をよく表現するモデルを用意できるならば、能率的な分析が可能。
  - （例）線形予測分析

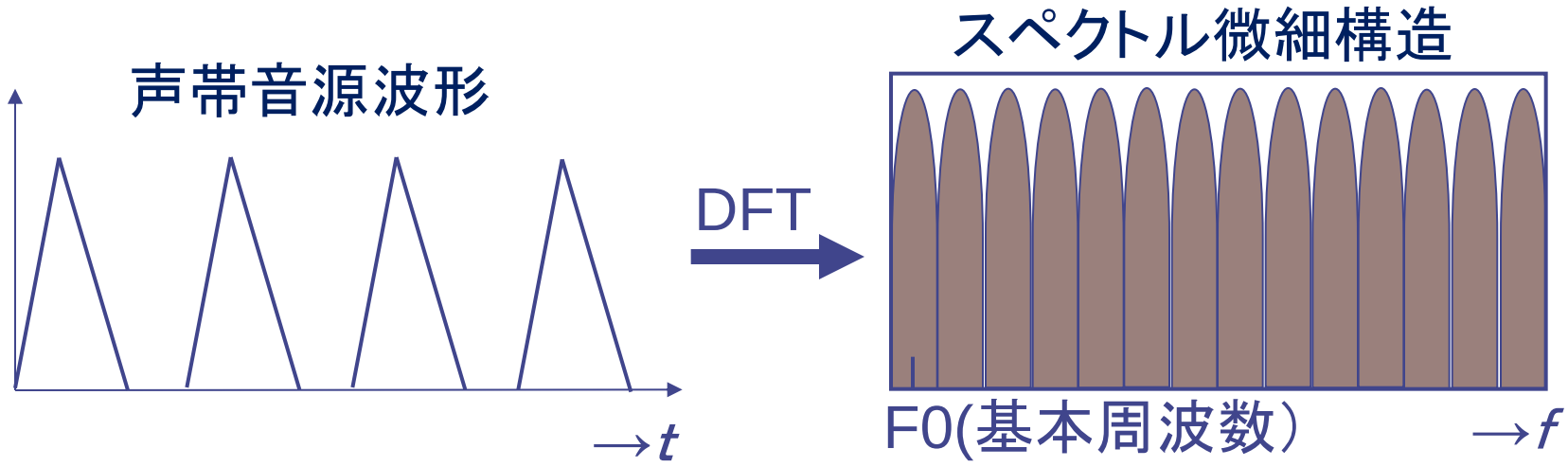
音声のスペクトル構造：声帯音源

- 数学的には…

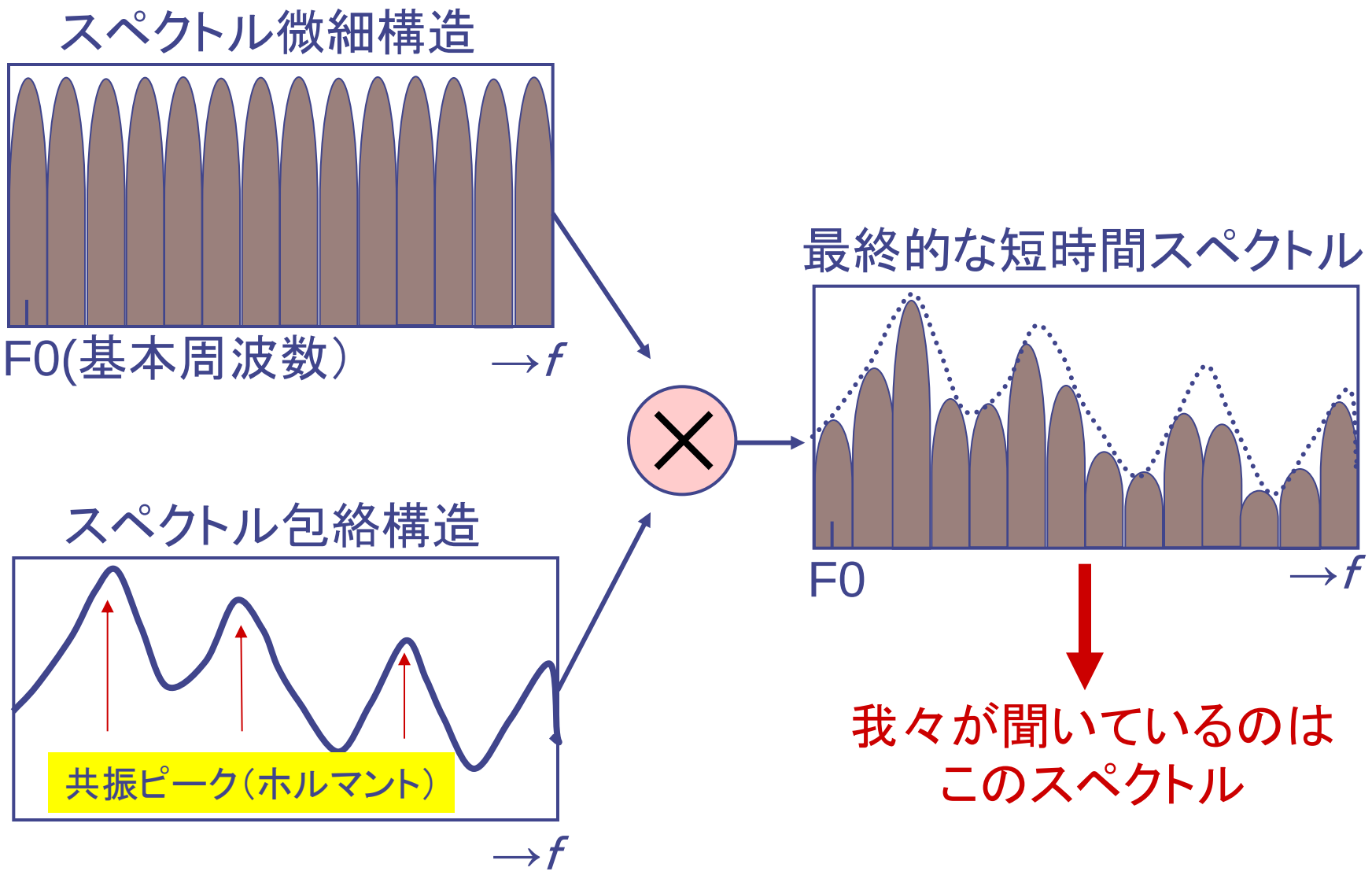
  $x(n)$ が周期信号である場合には、その周波数特性 $X(k)$ も周期関数となる。

- 一般に音声における声帯音源信号は…
  - 有声音の場合には「周期的三角波」

    ⇒　声帯音源の振幅スペクトルは周期的な微細構造を持つ。よって音声信号自体も周期的スペクトルとなる。



音声のスペクトル構造：分析例



音声スペクトルから得られる2種類の情報

■音声信号スペクトル

1. スペクトル微細構造

- 周期成分　⇒　声帯の振動に対応
- その人個人が持つ「声の高さ」

2. スペクトル包絡構造

- 声道・鼻腔における共振・反共振特性
  ⇒　各音韻ごとの違いに対応
- 音声認識処理などでは、この包絡情報に基づいて識別を行う.

両者が混合されて観測されるためその分離は不可能である

音声認識など音韻の識別には包絡情報のみが必要

⇒DFT分析だけでは不充分。更なる包絡抽出分析法が必要

スペクトル包絡の代表的抽出法

- ケプストラム法
  - モデルを仮定しないノンパラメトリック法の一種
  - 短時間スペクトル上において微細構造と包絡構造とを分ける。
- 線形予測法
  - 自己回帰モデルに基づくパラメトリック法
  - 声道における共振特性をモデリング

ケプストラムとは？

- ケプストラム（cepstrum）
  - 波形の短時間振幅スペクトルの対数の逆フーリエ変換として定義される。
  - ケプストラム領域では、微細構造と包絡構造に対応する各成分を容易に見分けることができる。

“Cepstrum”とは、「スペクトルを逆変換する」という意味を含めてspectrumをもじって作った造語である。
（Bogert, 1963）

ケプストラム算出手順

時間波形

フーリエ変換

短時間スペクトル

絶対値

振幅スペクトル

対数

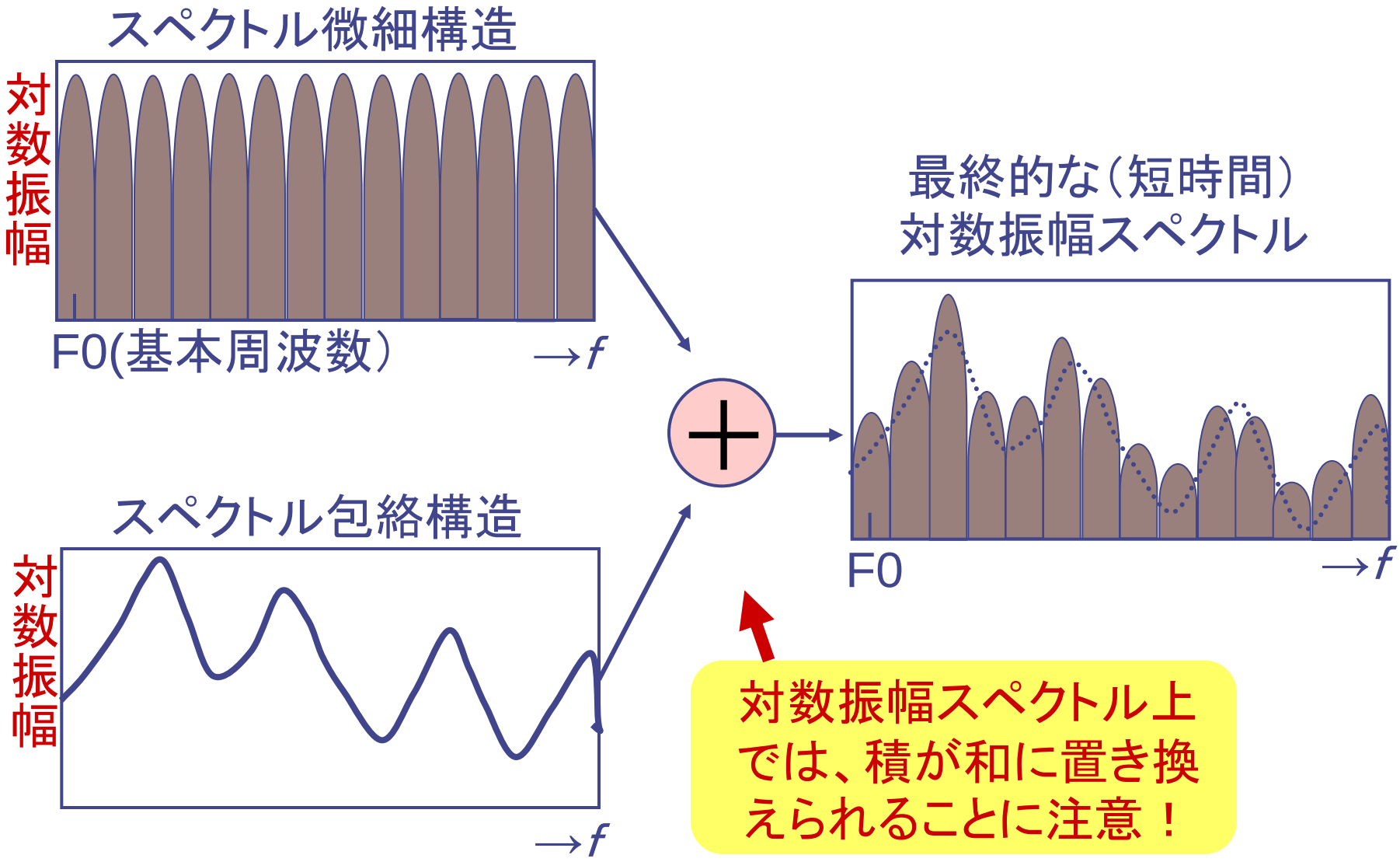
対数振幅スペクトル

逆フーリエ変換

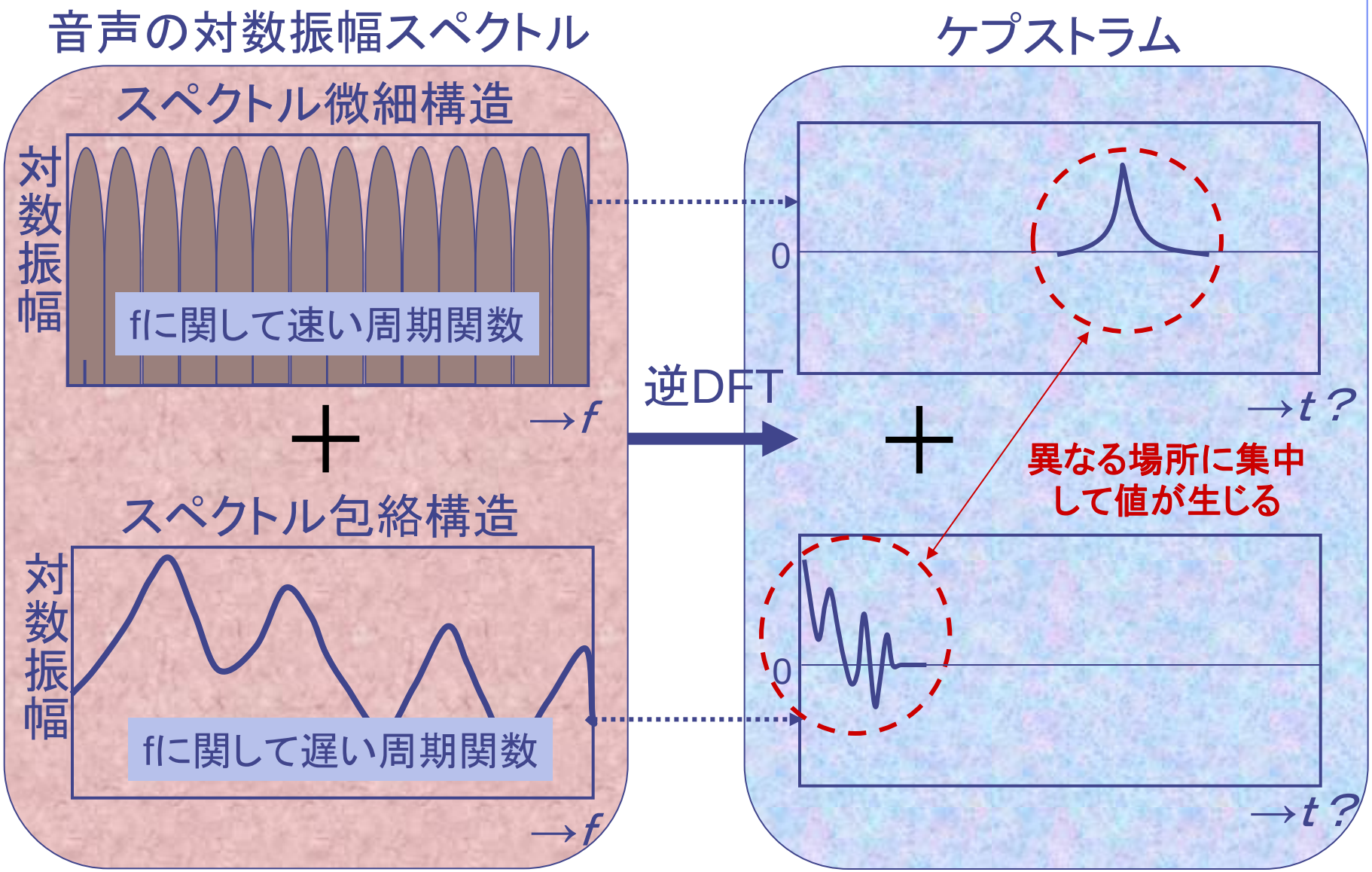
ケプストラム

フーリエ変換には、一般にDFTが用いられる。

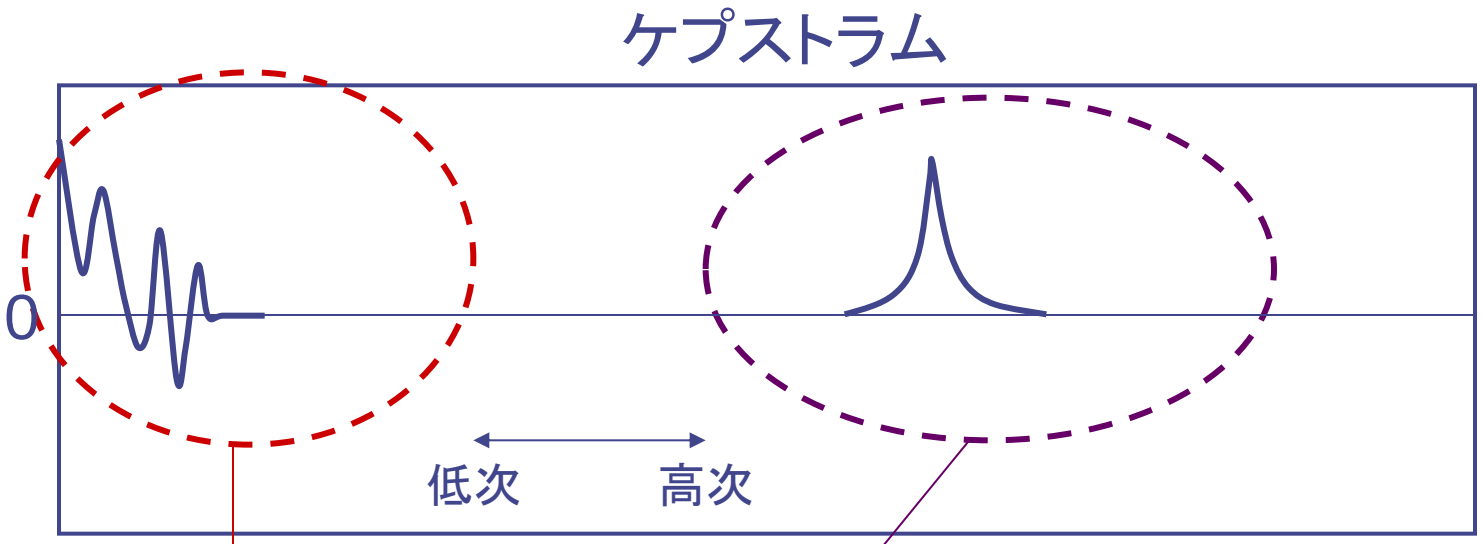
音声の「対数」振幅スペクトル構造



対数振幅スペクトルとケプストラム



典型的なケプストラムの構造



→ quefrency

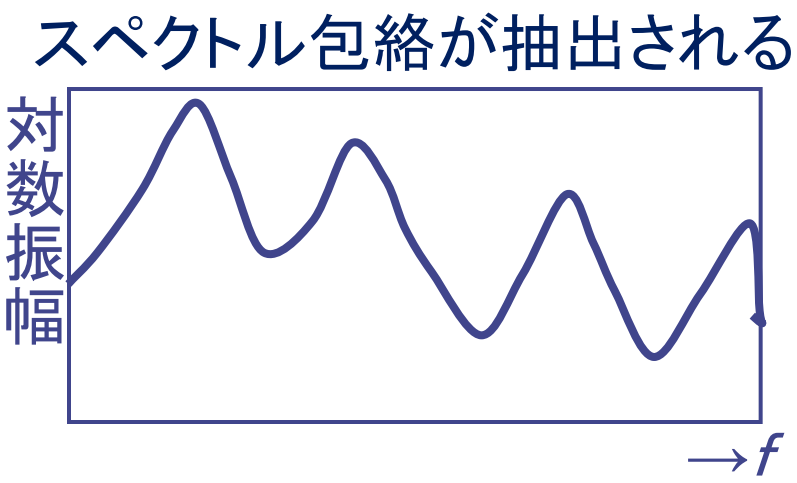
低次のケプストラムは
スペクトル包絡に対応

高次のケプストラムは
スペクトル微細構造に対応

それぞれ異なるケフレンシー位置に値が生じるので、容易に区別可能

単位は「時間」ではないが、周波数frequencyの逆数のようなものなので、それをもじって“quefrency”（ケフレンシー）と呼ばれる。

リフタリングによる包絡構造抽出



低次ケプストラムのみをケフレンシー領域で切り出す窓関数を“lifter”（リフター）*という。

＊周波数領域のfilterのもじり

ケプストラム処理の特徴

- スペクトル構造の分解
  - 対数を利用してスペクトル積を和に変換
  - ケフレンシー領域へ変換することにより、スペクトル包絡と周期的微細構造を区別可能にする。
  - 単純な窓かけ操作(リフター)により、包絡成分のみ(もしくは微細構造のみ)を抽出可能

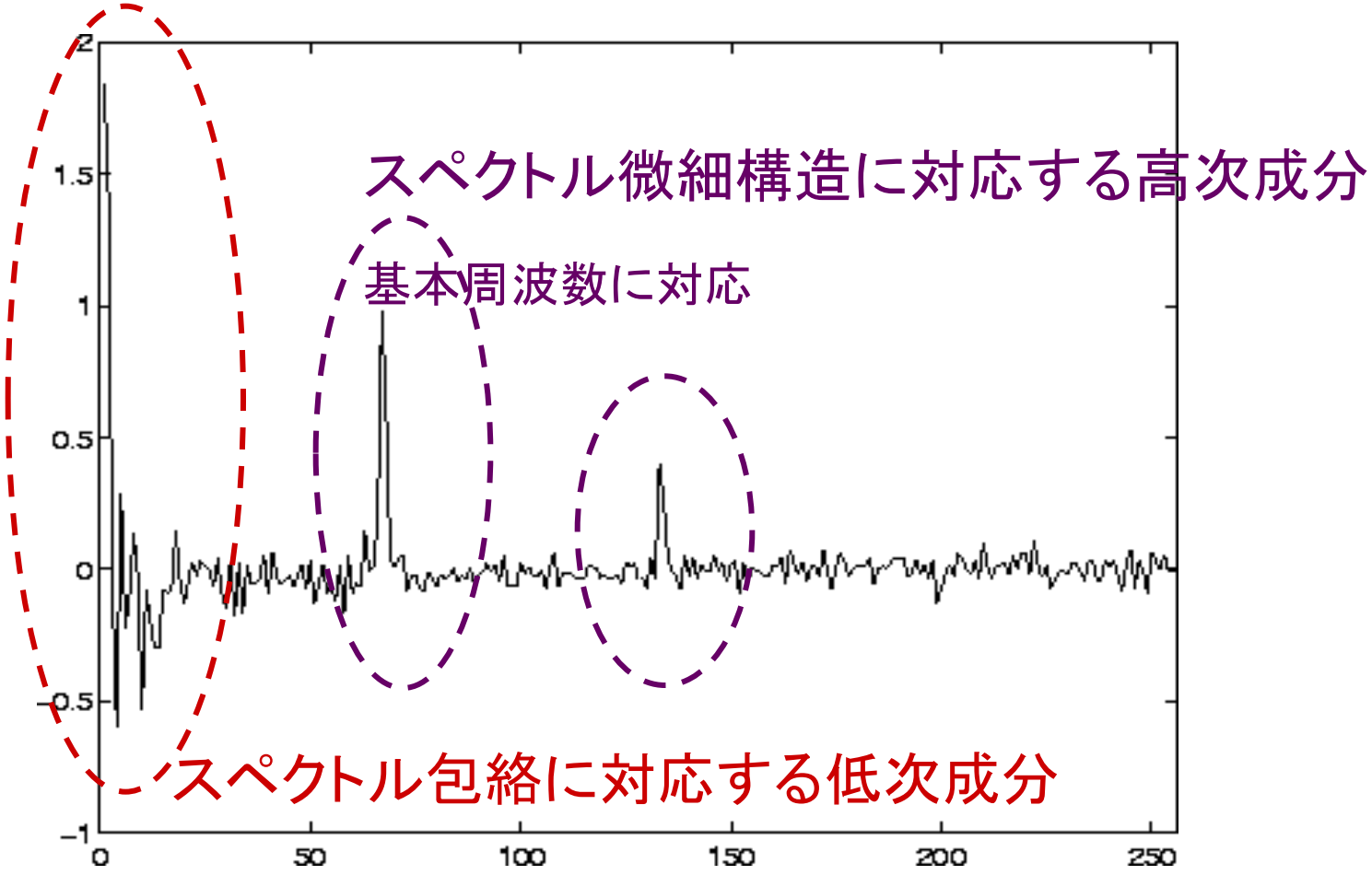
- 少ない演算量
  - スペクトル包絡成分を抽出するのに必要な演算

    [対数演算＋逆DFT＋リフタリング＋DFT]

    ⇒　非常に少ない演算量で抽出可能

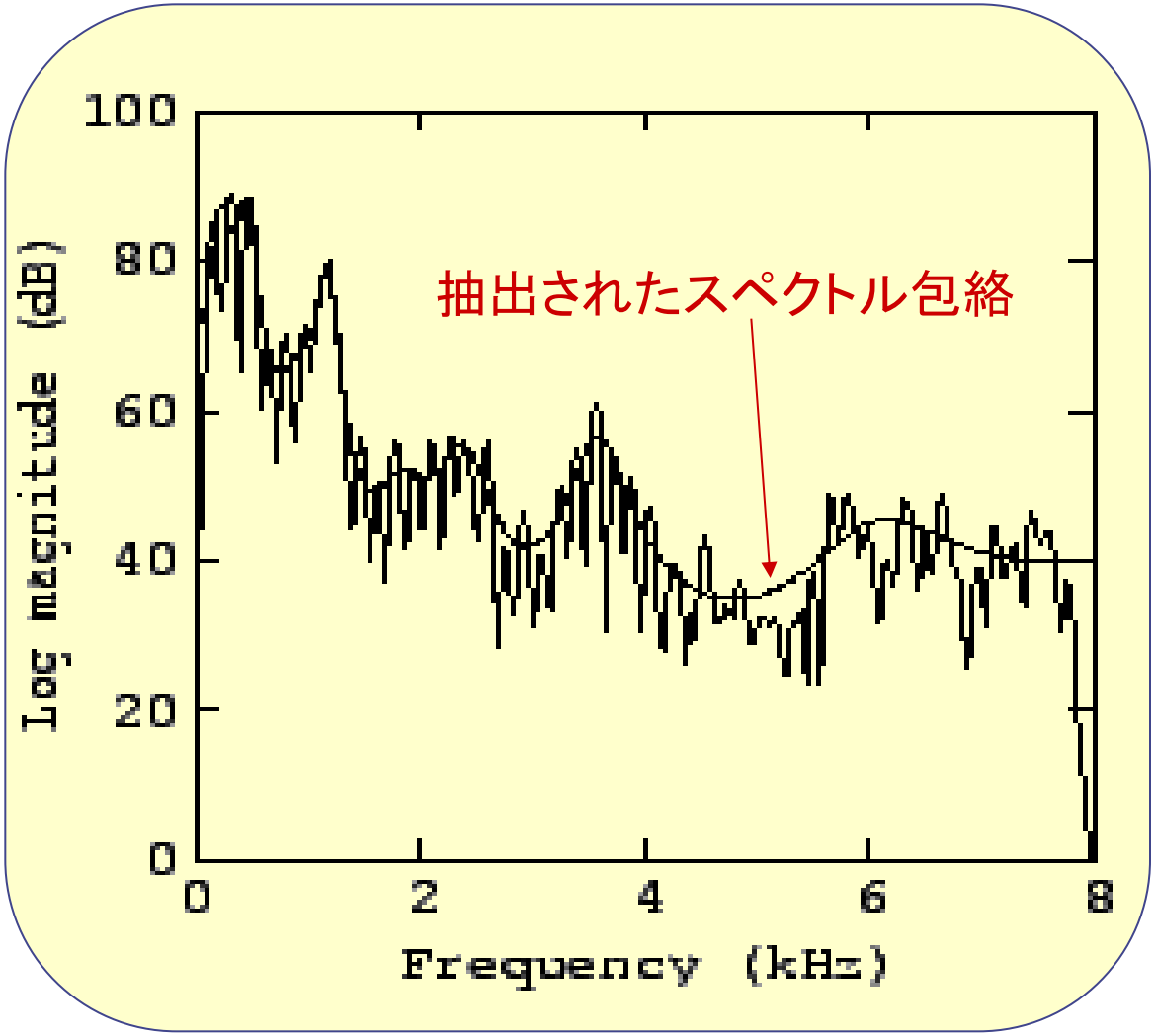
ケプストラム例

母音のケプストラムの典型例

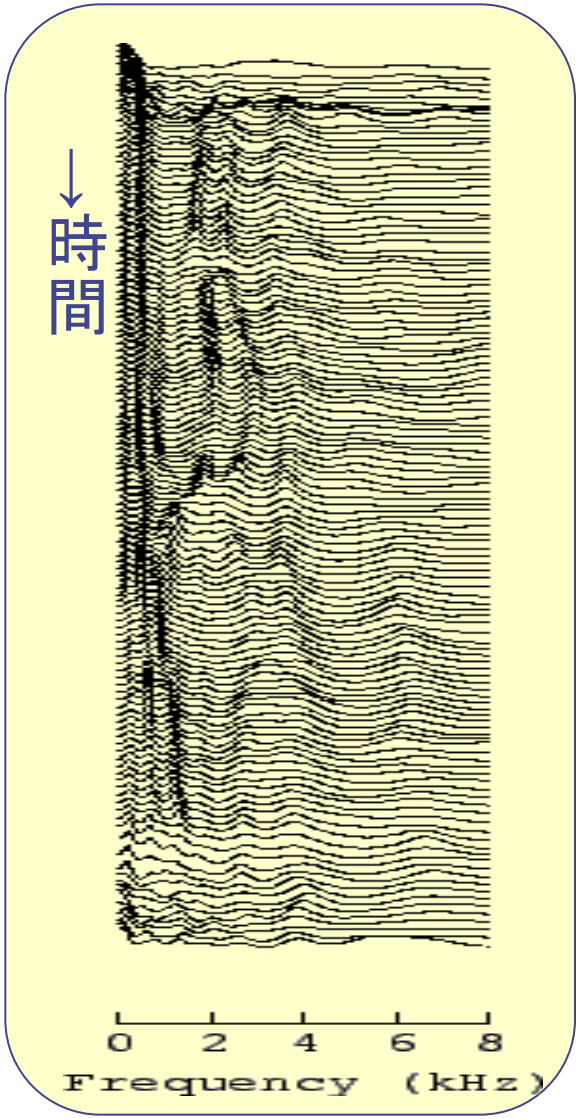


ケプストラムによるスペクトル包絡

短時間スペクトル包絡の例



スペクトル包絡の時間遷移



ケプストラム分析のまとめ

- 長所
  - 比較的単純な操作でスペクトル包絡抽出可能
  - 高次ケプストラムも使用すれば基本周波数も抽出可能
- 問題点
  - リフタリングのカットオフ位置をどのようにして決めるか？
  - 抽出されたスペクトル包絡において、ホルマント共振があまり強く表示されない。

人間の聴覚系では共振点をより聞いていると言われている

⇒声道での共振をモデルにしたパラメトリック分析が有効