

歌声・音声分析合成のための F0適応多重フレーム統合分析に基づく スペクトル包絡と群遅延の推定法

中野 倫靖, 後藤 真孝
(産業技術総合研究所)

2012年8月9日

第96回 音楽情報科学研究会 (SIGMUS夏のシンポジウム)

研究目的

□ 声（歌声・話声）と音（楽器音）の高性能な分析合成



■ 高精度な分析

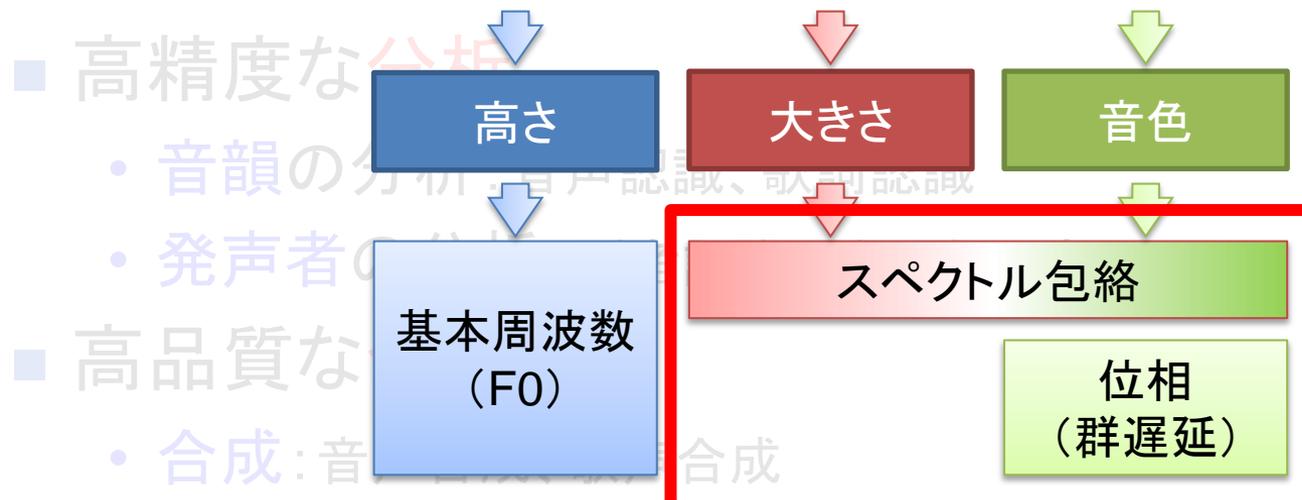
- 音韻の分析：音声認識、歌詞認識
- 発声者の分析：話者認識、歌手名同定

■ 高品質な合成

- 合成：音声合成、歌声合成
- 変換：声質変換、韻律変換

研究目的

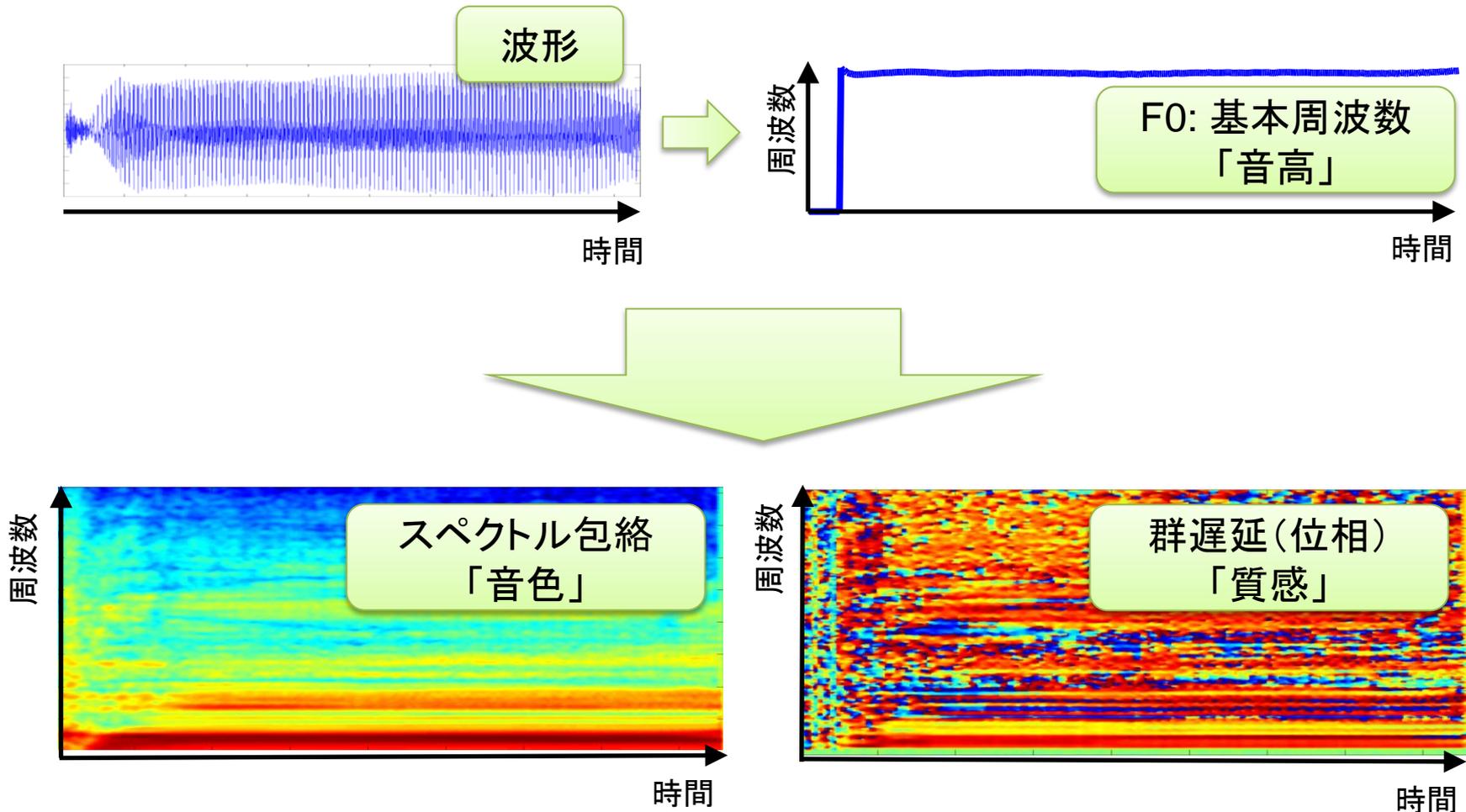
声(歌声・話声)と音(楽器音)の高性能な分析合成



- 高精度な分析
 - ・ 音韻の分析
 - ・ 発声者の分析
- 高品質な合成
 - ・ 合成: 音質変換
 - ・ 変換: 声質変換、韻律変換

本研究で対象

提案手法の入出力

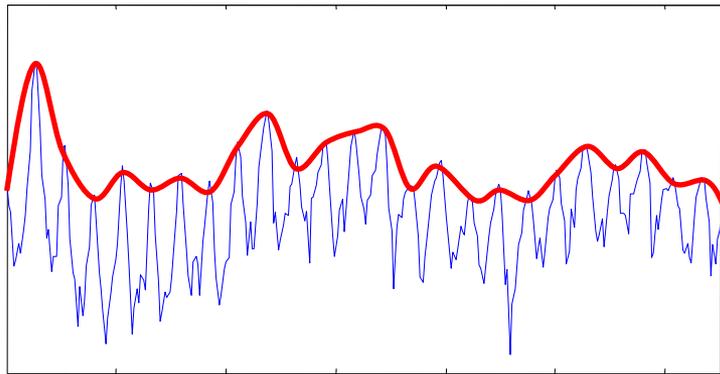


F0適応多重フレーム統合分析に基づく スペクトル包絡の推定

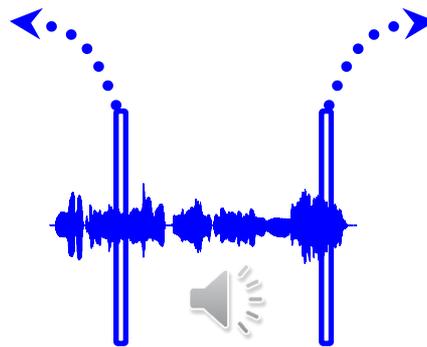
スペクトル包絡とは

□ 振幅スペクトルの概形

- 周波数帯域毎の音の大きさ
- 声道特性(音韻の違い)や音色の違いを反映

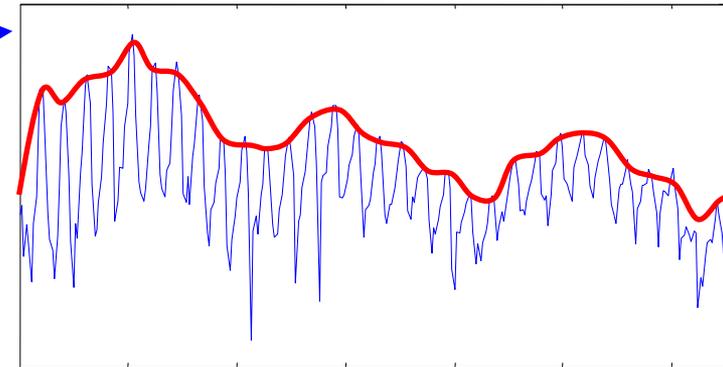


「い」の
振幅スペクトルと
スペクトル包絡



歌声の波形
(信号)の例

「吞めといわれて陽気に吞んだ」



「あ」の
振幅スペクトルと
スペクトル包絡

スペクトル包絡の推定方法(従来)

- 振幅スペクトルに基づく方法：フィッティング(最尤推定)
[Itakura et. al.,1968, Atal et. al.,1971] [今井1979]

- LPC, ケプストラム(改良ケプストラム法)

その他: Phase vocoder

- F0推定を伴う方法

- [Moulines et. al.,1999] [Kawahara et. al.,1999] [Kawahara et. al.,2008]
 - PSOLA分析, STRAIGHT, TANDEM-STRAIGHT

その他: 正弦波モデル, GMM

- 複数フレームから推定する方法

- [Akamine et. al.,1998, Shiga et. al.,2003, Toda et. al.,2008, Fujihara et. al.,2009]
 - 同じ音素で異なる音高の別フレームの情報を活用

スペクトル包絡の推定方法(従来)

- 振幅スペクトルに基づく方法：フィッティング(最尤推定)
[Itakura et. al.,1968, Atal et. al.,1971] [今井1979]

- LPC, ケプストラム(改良ケプストラム法)

その他: Phase vocoder

- F0推定を伴う方法

- [Moulines et. al.,1999] [Kawahara et. al.,1999] [Kawahara et. al.,2008]
 - PSOLA分析, STRAIGHT, TANDEM-STRAIGHT

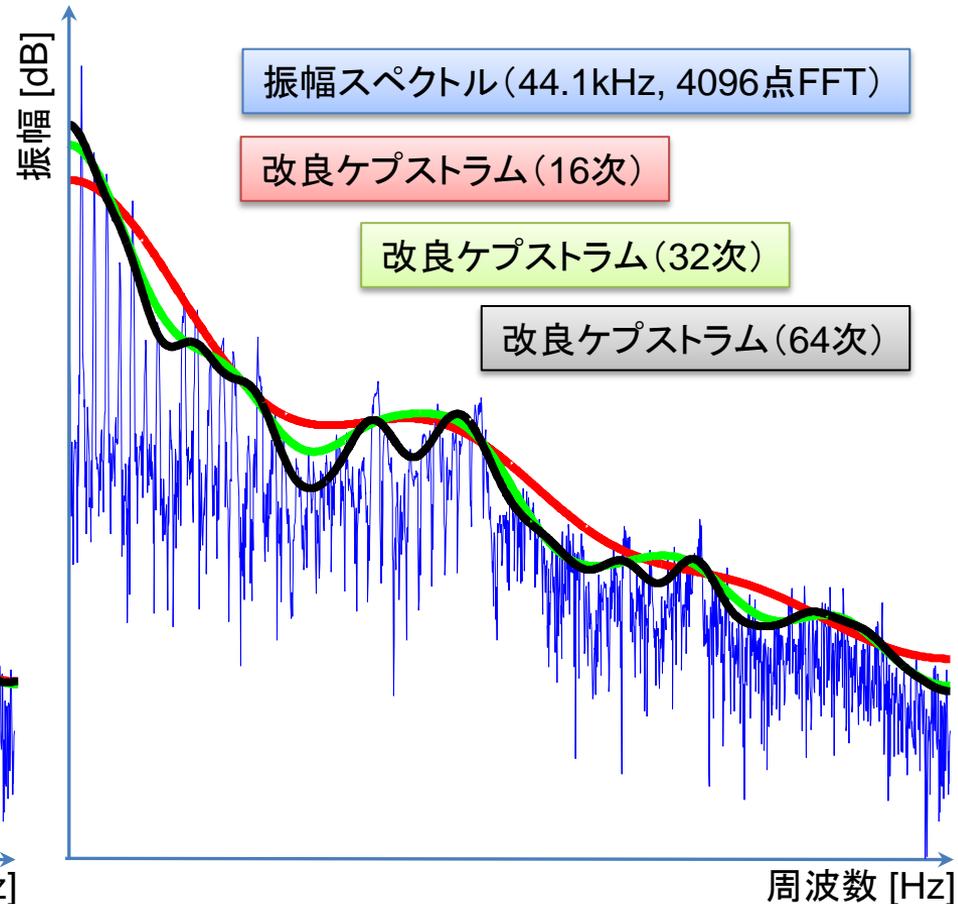
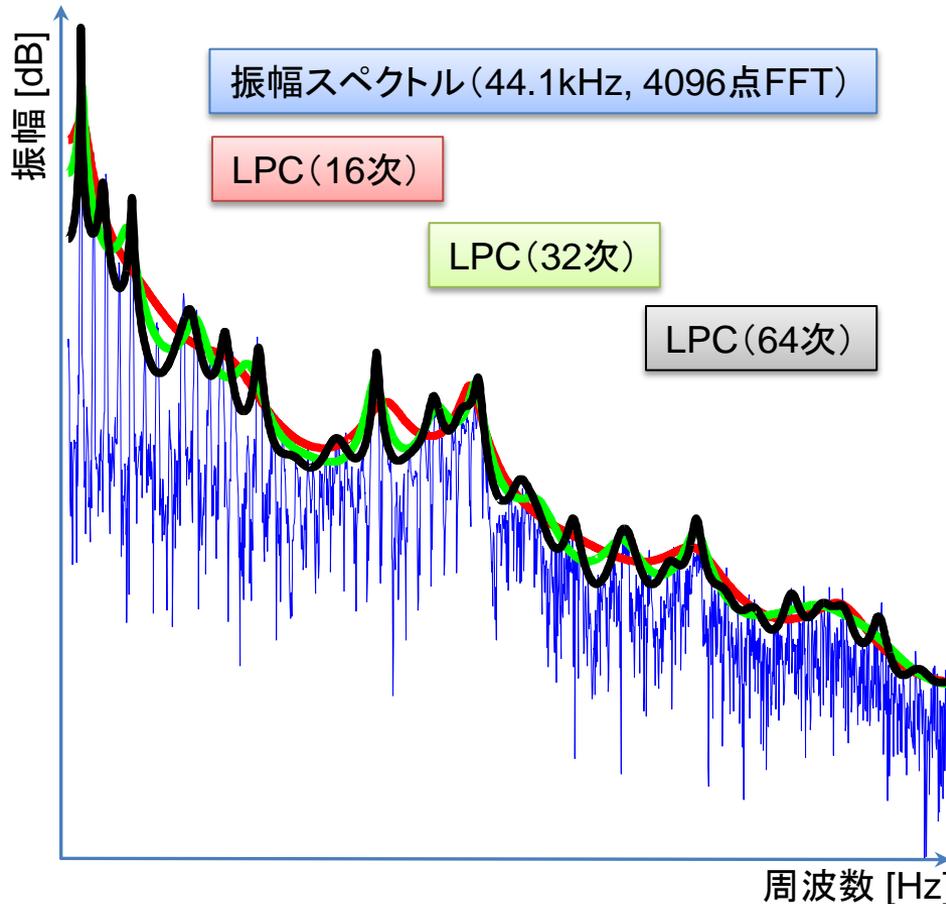
その他: 正弦波モデル, GMM

- 複数フレームから推定する方法

- [Akamine et. al.,1998, Shiga et. al.,2003, Toda et. al.,2008, Fujihara et. al.,2009]
 - 同じ音素で異なる音高の別フレームの情報を活用

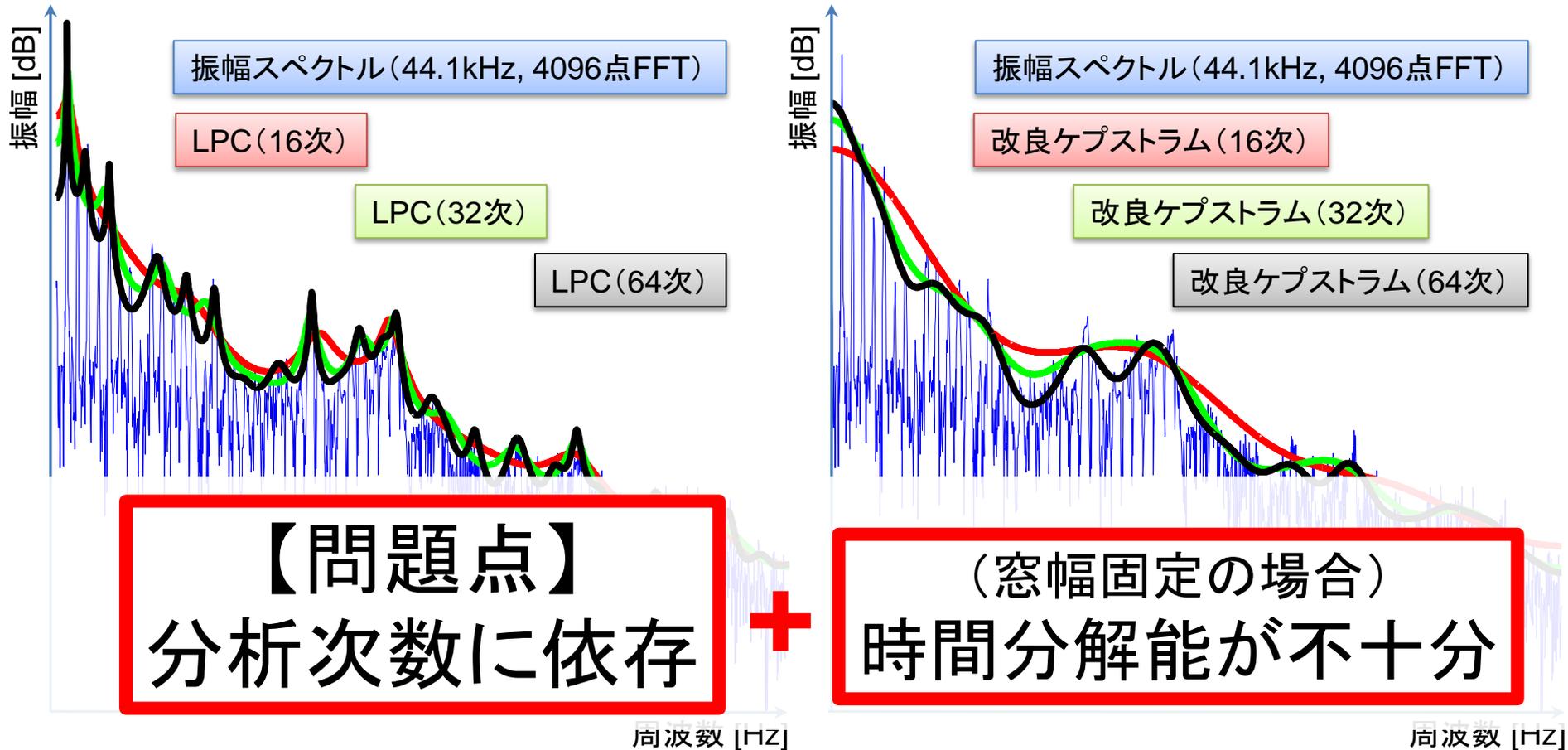
振幅スペクトルに基づく方法 (LPC, 改良ケプストラム)

□ 次数を決めてフィッティングする



振幅スペクトルに基づく方法 (LPC, 改良ケプストラム)

□ 次数を決めてフィッティングする



スペクトル包絡の推定方法(従来)

- 振幅スペクトルに基づく方法：フィッティング(最尤推定)
[Itakura et. al.,1968, Atal et. al.,1971] [今井1979]

- LPC, ケプストラム(改良ケプストラム法)

その他: Phase vocoder

- F0推定を伴う方法

- PSOLA分析, STRAIGHT, TANDEM-STRAIGHT
[Moulines et. al.,1999] [Kawahara et. al.,1999] [Kawahara et. al.,2008]

その他: 正弦波モデル, GMM

- 複数フレームから推定する方法

- 同じ音素で異なる音高の別フレームの情報を活用
[Akamine et. al.,1998, Shiga et. al.,2003, Toda et. al.,2008, Fujihara et. al.,2009]

PSOLA (ピッチマーク) 分析

[Moulines et. al., 1999]

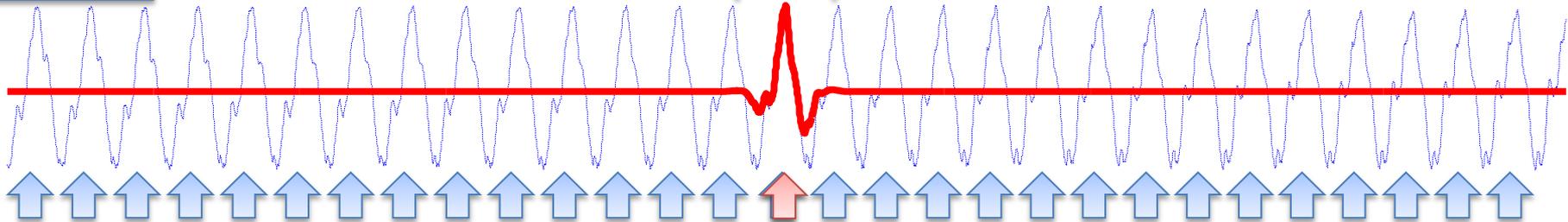
□ 基本周期に同期した分析

■ ピッチマーク

波形の駆動点かつ分析時刻を示す時刻情報

から **2周期** 切り出して包絡を近似

歌声波形



ピッチマーク

2周期の長さの窓関数を掛けて切り出す = **単位波形**

PSOLA (ピッチマーク) 分析

[Moulines et. al.,1999]

基本周期に同期した分析

■ ピッチマーク

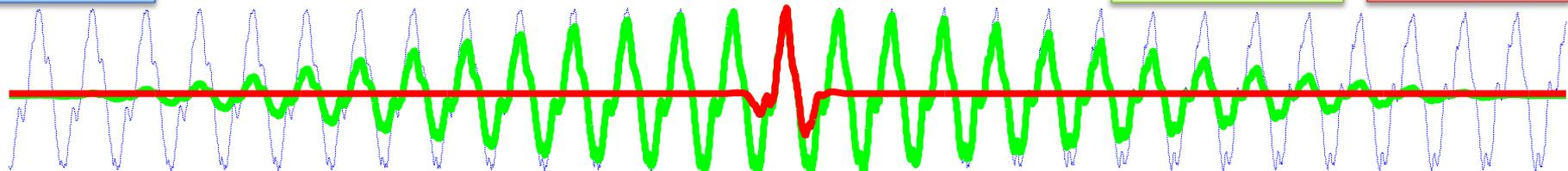
波形の駆動点かつ分析時刻を示す時刻情報

から2周期切り出して包絡を近似

歌声波形

ハニング窓

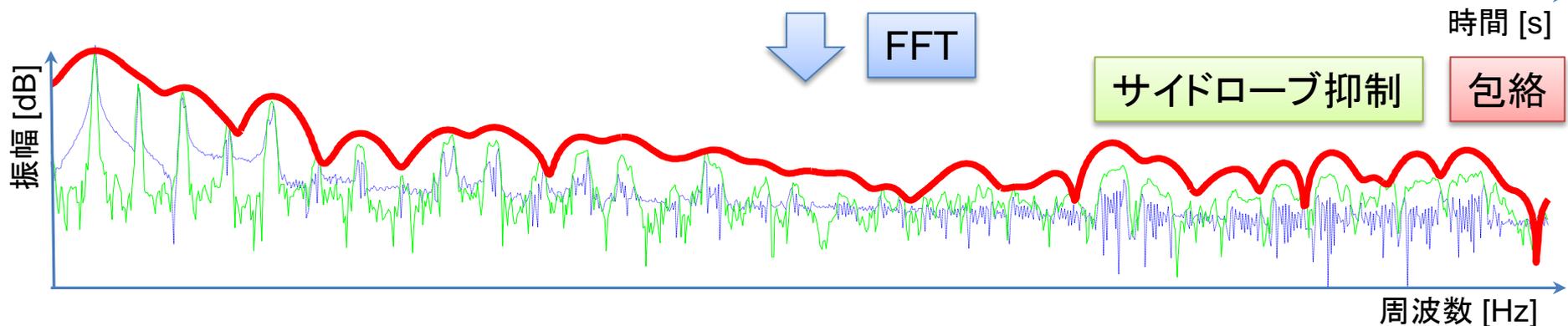
単位波形



FFT

サイドローブ抑制

包絡



PSOLA (ピッチマーク) 分析

[Moulines et. al., 1999]

基本周期に同期した分析

- **ピッチマーク** 波形の駆動点かつ分析時刻を示す時刻情報

から**2周期**切り出して包絡を近似

歌声波形

ハニング窓

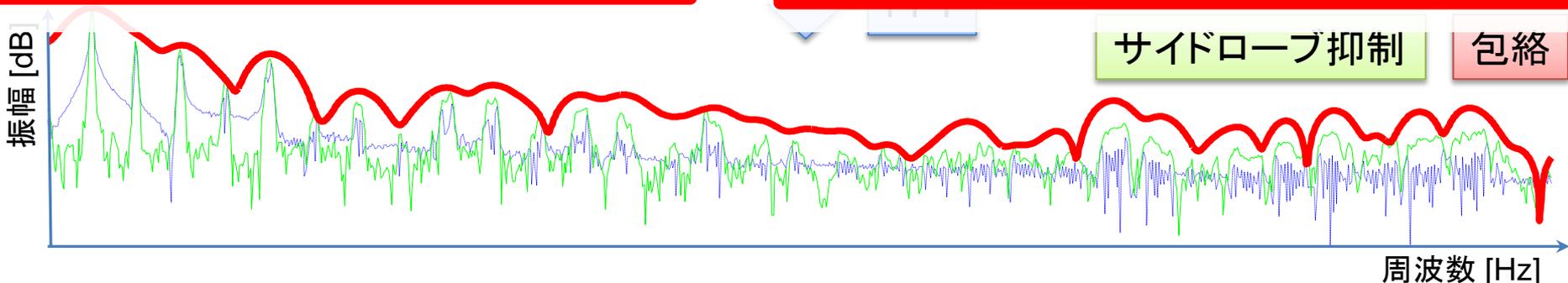
単位波形

【問題点】

ピッチマークが必要

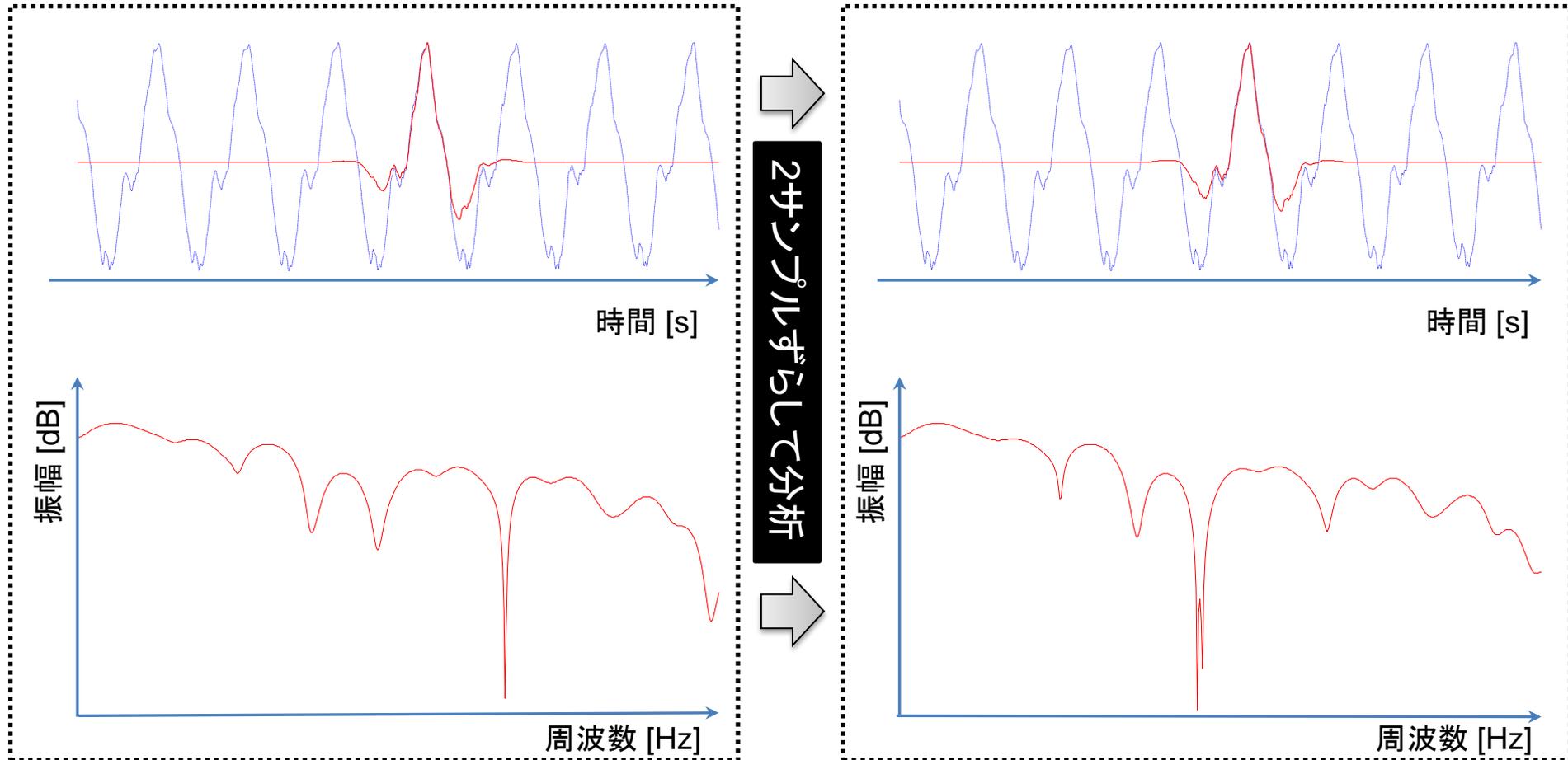
+

**真の包絡が得られない
(次のスライド)**



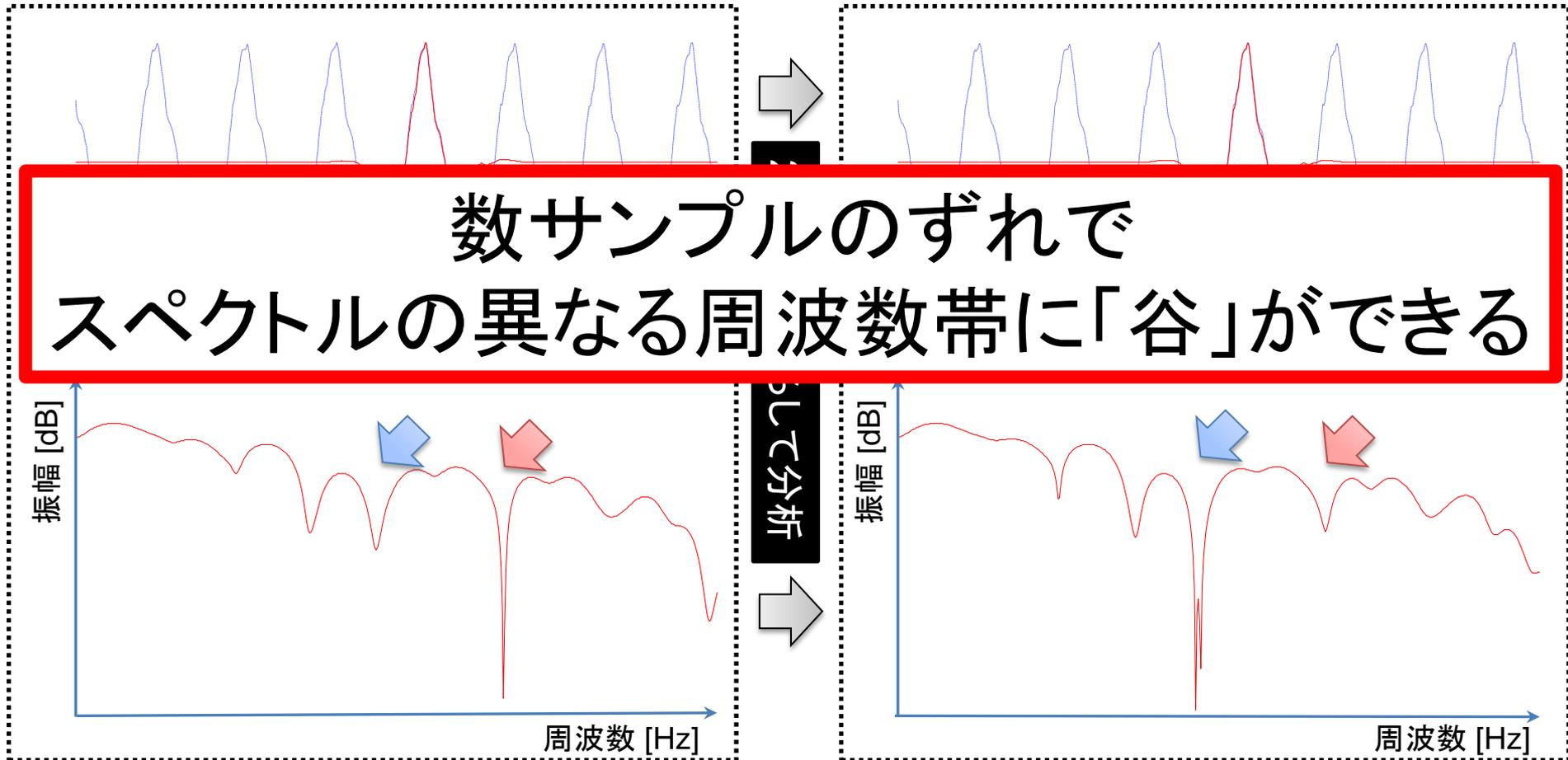
ピッチマーク分析の問題点

- 単一のピッチマークから真の包絡線が得られない



ピッチマーク分析の問題点

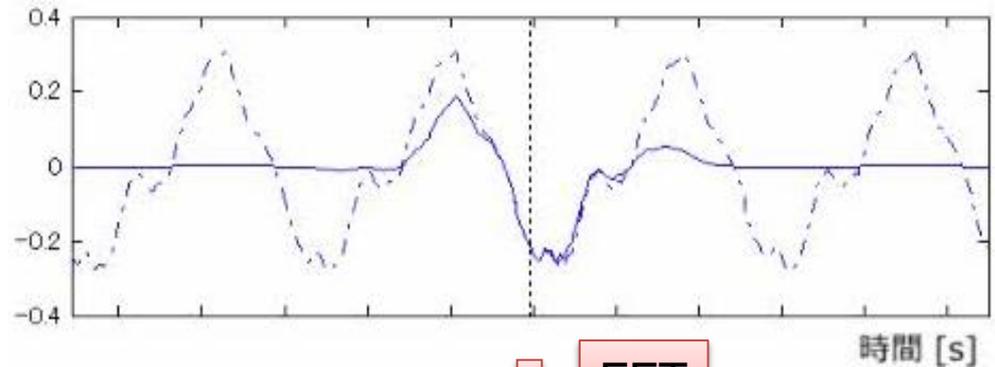
- 単一のピッチマークから真の包絡線が得られない



ピッチマーク分析の問題点

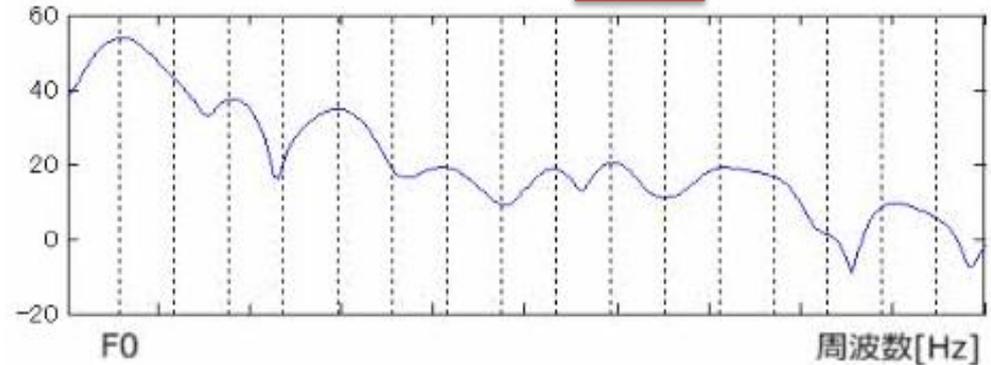
- 単一のピッチマークから真の包絡が得られない

分析時刻と
その振幅スペクトル
(スペクトル包絡)



FFT

どこで分析しても
谷が発生



スペクトル包絡の推定方法(従来)

- 振幅スペクトルに基づく方法：フィッティング(最尤推定)
[Itakura et. al.,1968, Atal et. al.,1971] [今井1979]

- LPC, ケプストラム(改良ケプストラム法)

その他: Phase vocoder

- F0推定を伴う方法

- PSOLA分析, [Moulines et. al.,1999] STRAIGHT, [Kawahara et. al.,1999] TANDEM-STRAIGHT, [Kawahara et. al.,2008]

その他: 正弦波モデル, GMM

- 複数フレームから推定する方法

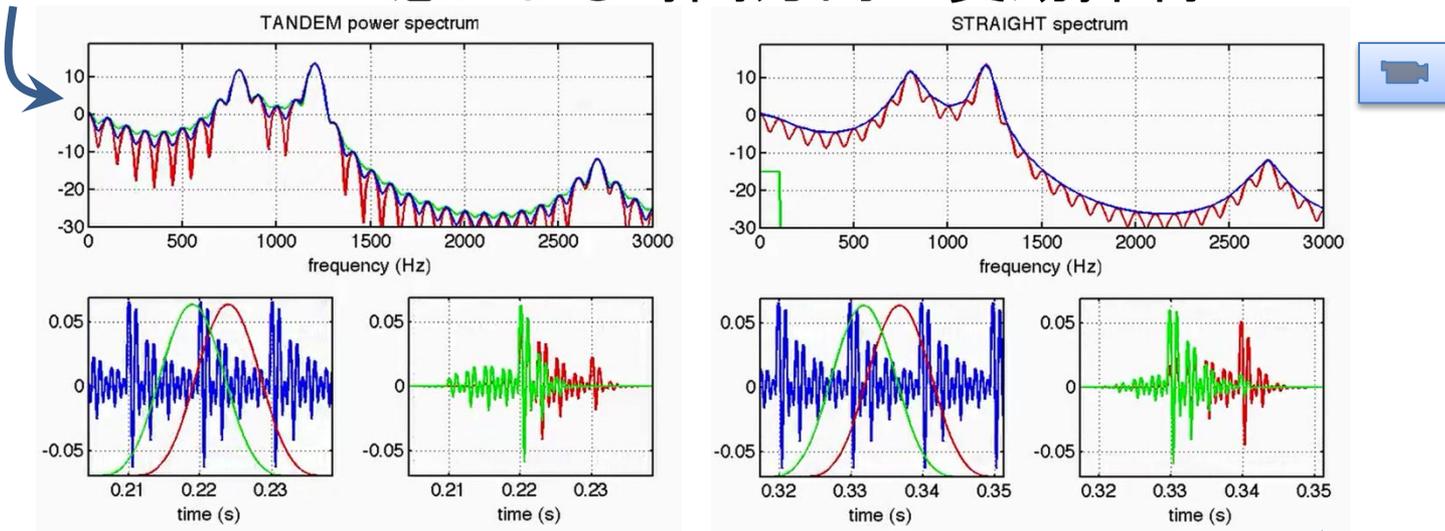
- 同じ音素で異なる音高の別フレームの情報を活用
[Akamine et. al.,1998, Shiga et. al.,2003, Toda et. al.,2008, Fujihara et. al.,2009]

F0推定を伴う方法 (STRAIGHT, TANDEM-STRAIGHT) [Kawahara et. al.,1999] [Kawahara et. al.,2008]

□ F0に基づく二段階の処理

1. F0に基づく窓を用いた分析 (FFT)

- TANDEM窓による時間方向の変動抑制



2. F0に基づく周波数スペクトルの平滑化

ムービー引用先URL: <http://www.wakayama-u.ac.jp/~kawahara/SIGMUS94kawahara/SIGMUS94files/index.html>

F0推定を伴う方法 (STRAIGHT, TANDEM-STRAIGHT)

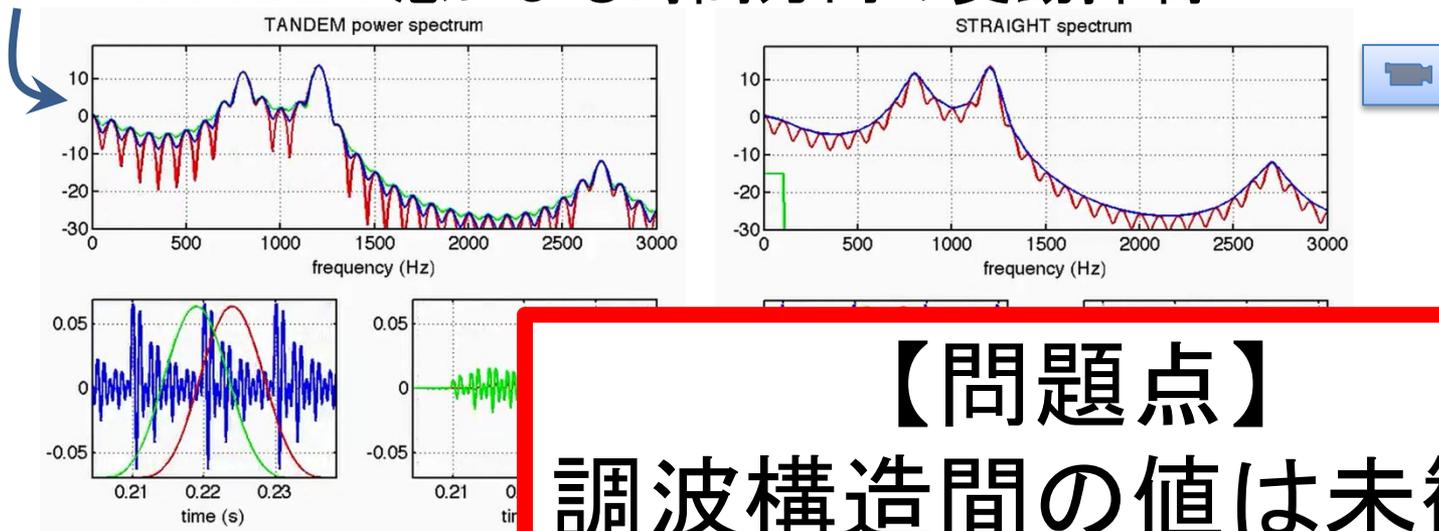
[Kawahara et. al.,1999]

[Kawahara et. al.,2008]

□ F0に基づく二段階の処理

1. F0に基づく窓を用いた分析 (FFT)

- TANDEM窓による時間方向の変動抑制



【問題点】
 調波構造間の値は未観測
 (平滑化によって推定)

2. F0に基づく周波数 (平滑化によって推定)

ムービー引用先URL: <http://www.wakayama-u.ac.jp/~kawahara/SIGMUS94kawahara/SIGMUS94files/index.html>

スペクトル包絡の推定方法(従来)

- 振幅スペクトルに基づく方法：フィッティング(最尤推定)
[Itakura et. al.,1968, Atal et. al.,1971] [今井1979]

- LPC, ケプストラム(改良ケプストラム法)

その他: Phase vocoder

- F0推定を伴う方法

- [Moulines et. al.,1999] [Kawahara et. al.,1999] [Kawahara et. al.,2008]
 - PSOLA分析, STRAIGHT, TANDEM-STRAIGHT

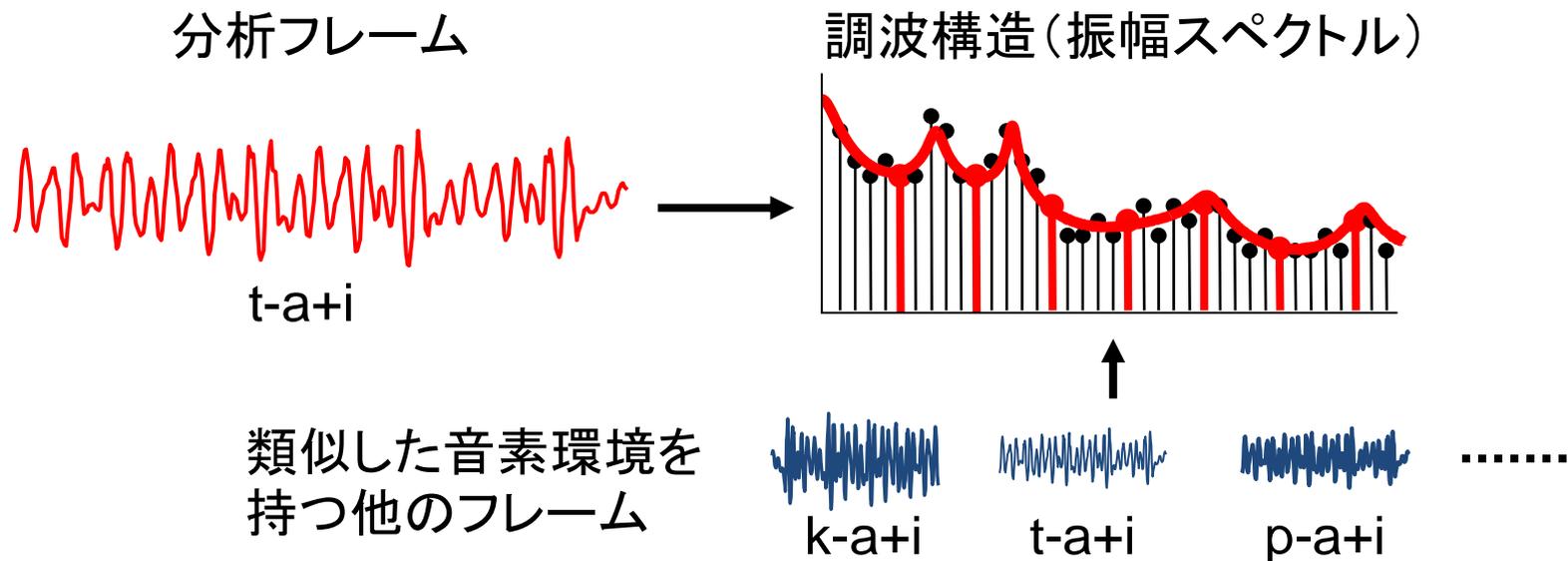
その他: 正弦波モデル, GMM

- 複数フレームから推定する方法

- [Akamine et. al.,1998, Shiga et. al.,2003, Toda et. al.,2008, Fujihara et. al.,2009]
 - 同じ音素で異なる音高の別フレームの情報を活用

複数フレーム分析：複数フレームから包絡を推定

□ 調波構造間の値を他の音高での値から推定



【問題点】

音素ラベルが必要

+

**コンテキスト(歌い方等)の変動で
過剰な平滑化の可能性**

従来の問題点(1)

- 時間分解能が低い (窓幅固定の場合)
 - LPC, ケプストラム, 複数フレーム分析
- 推定形状が分析次数に依存
 - LPC, ケプストラム
- 単位波形の分析では不十分 (スペクトルの谷ができる)
 - PSOLA分析
- 調波構造間の値が観測できない
 - STRAIGHT, TANDEM-STRAIGHT
- コンテキストの変動へ対処が必要
 - 複数フレーム分析

従来の問題点(1)

- 時間分解能が低い (窓幅固定の場合)
 - LPC, ケプストラム, 複数フレーム分析
- 推定形状が分析次数に依存
 - LPC, ケプストラム
- 単位波形の分析では不十分 (スペクトルの谷ができる)
 - PSOLA分析
- 調波構造間の値が観測できない
 - STRAIGHT, TANDEM-STRAIGHT
- コンテキストの変動
 - 複数フレーム分析

【共通の問題点】
真の包絡を得るのが難しい

従来の問題点(2)

□ ピッチマークが必要

- PSOLA分析

□ 音素ラベルが必要

- 複数フレーズ

【ピッチマークの問題点】
本質的に真の値が得られない

【共通の問題点】
手作業を伴う事前の修正作業が必要
(自動で100%の精度を得ることが難しい)

提案手法：F0適応多重フレーム統合分析法

□ 特長

真の包絡の推定

ラベルを使わない

□ 二つの分析方法の組み合わせで実現

■ F0適応分析

真の包絡の推定

- 基本周期の2.0倍の長さの窓で分析（従来と同様）
- 全サンプリング点で分析する：速い変化を追従する

■ 多重フレーム統合分析

真の包絡の推定

ラベルを使わない

- 近傍の複数フレームを統合する
 - 単一のピッチマークよりも頑健（スペクトルの谷が発生しない）
 - ピッチマークや音素ラベルが不要

提案手法：F0適応多重フレーム統合分析法

□ 特長

真の包絡の推定

ラベルを使わない

□ 二つの分析方法の組み合わせで実現

■ F0適応分析

真の包絡の推定

- 基本周期の2.0倍の長さの窓で分析（従来と同様）
- 全サンプリング点で分析する：速い変化を追従する

■ 多重フレーム統合分析

真の包絡の推定

ラベルを使わない

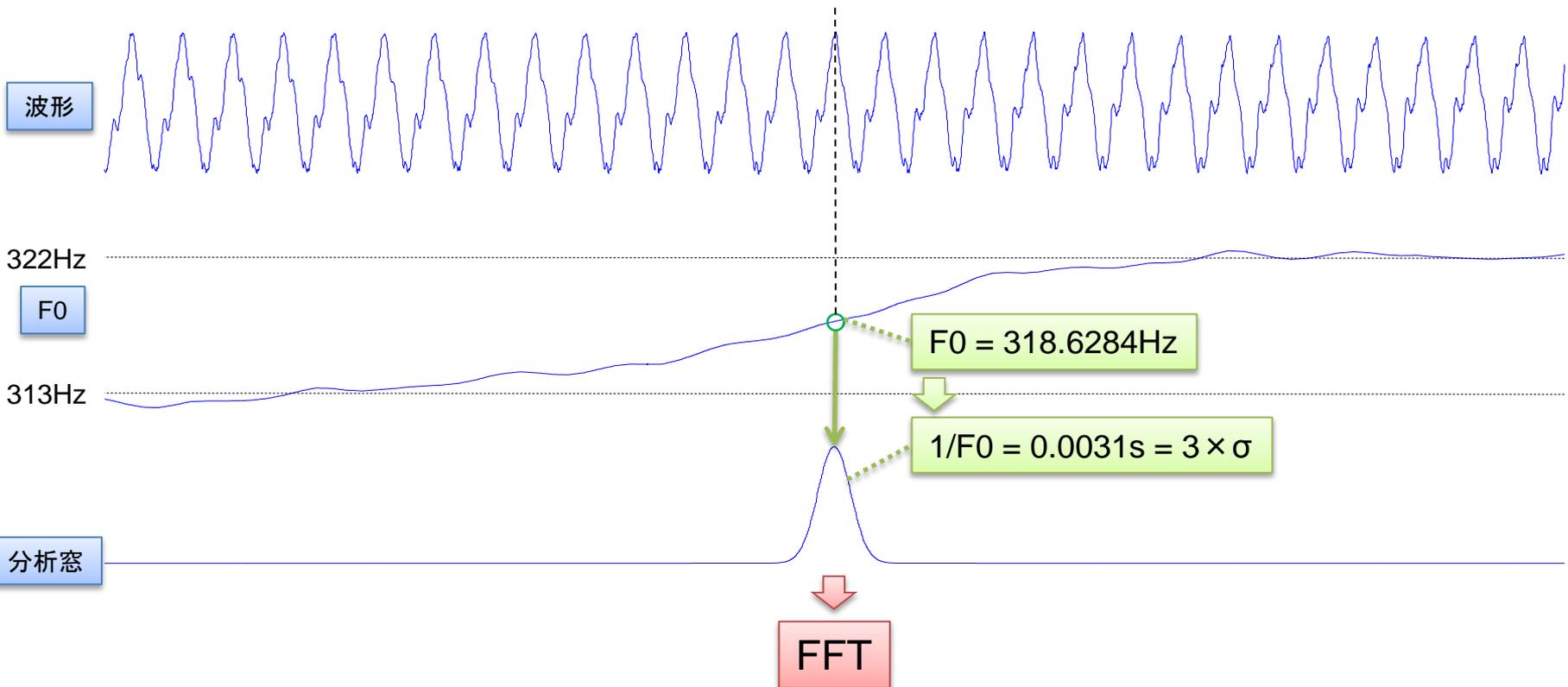
- 近傍の複数フレームを統合する
 - 単一のピッチマークよりも頑健（スペクトルの谷が発生しない）
 - ピッチマークや音素ラベルが不要

F0適応分析

- F0に適応した長さの短い窓で
全時間(全サンプリング点)に対してFFT

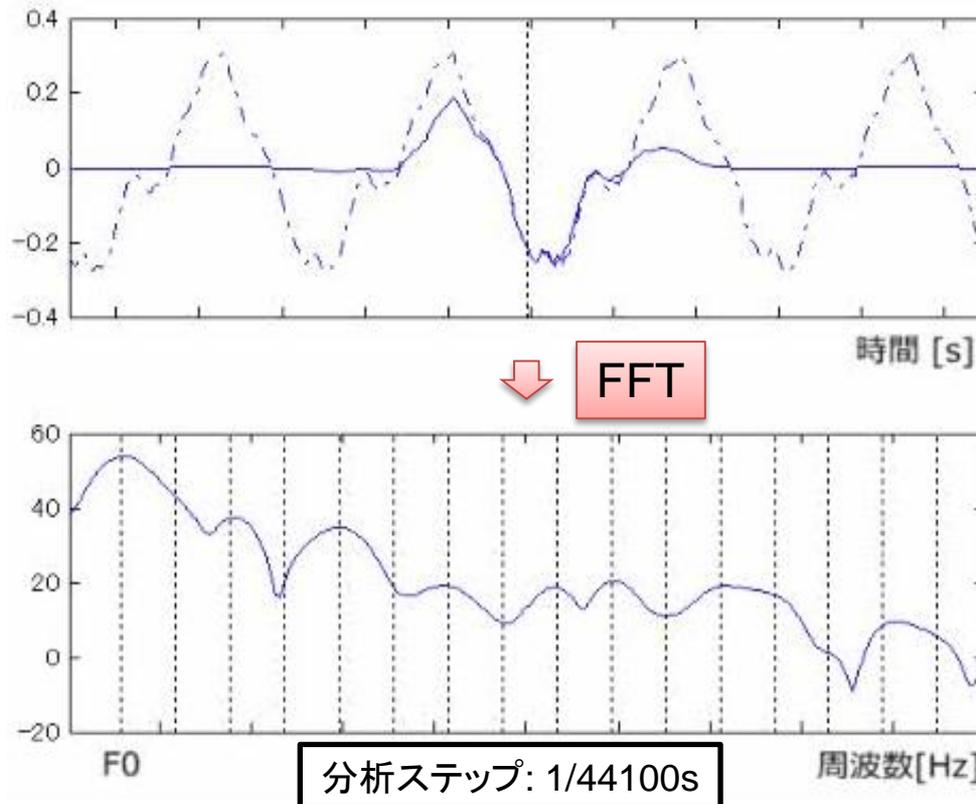
F0適応分析:F0に適応した長さの短い窓

- F0に適応した長さの短い窓で
全時間(全サンプリング点)に対してFFT

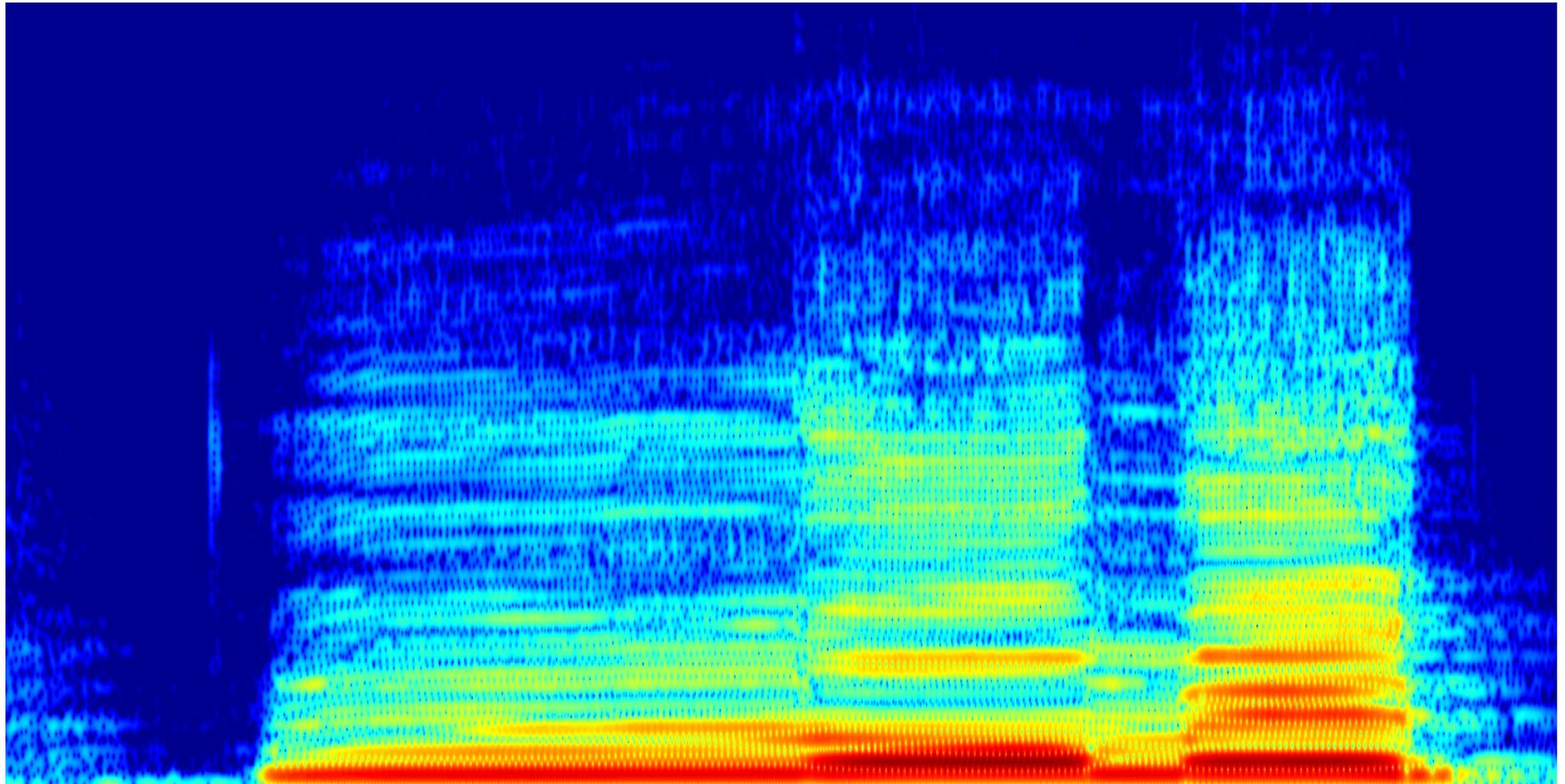
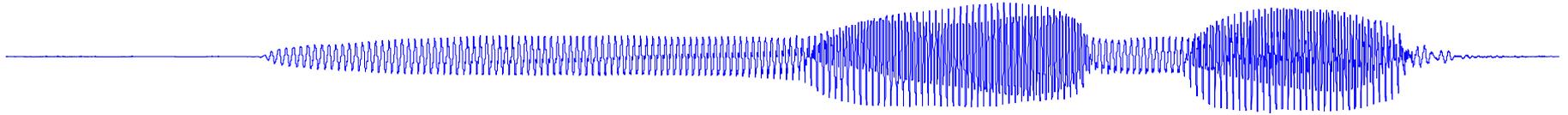


F0適応分析：全時間に対してFFT

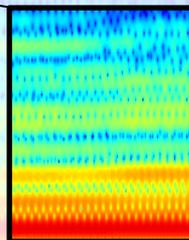
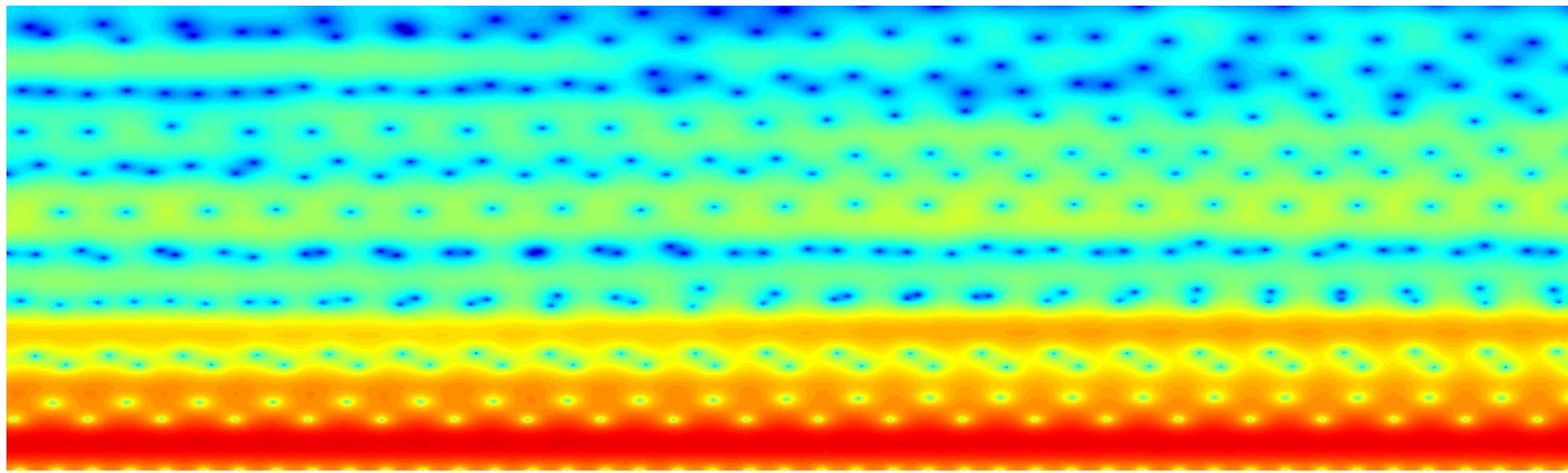
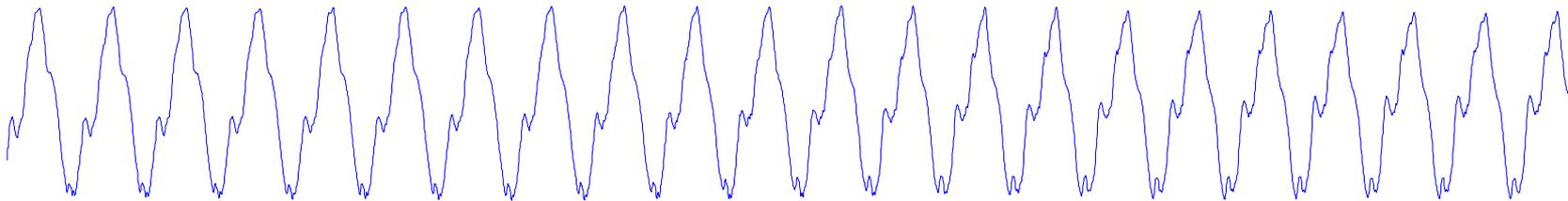
- F0に適応した長さの短い窓で
全時間(全サンプリング点)に対してFFT



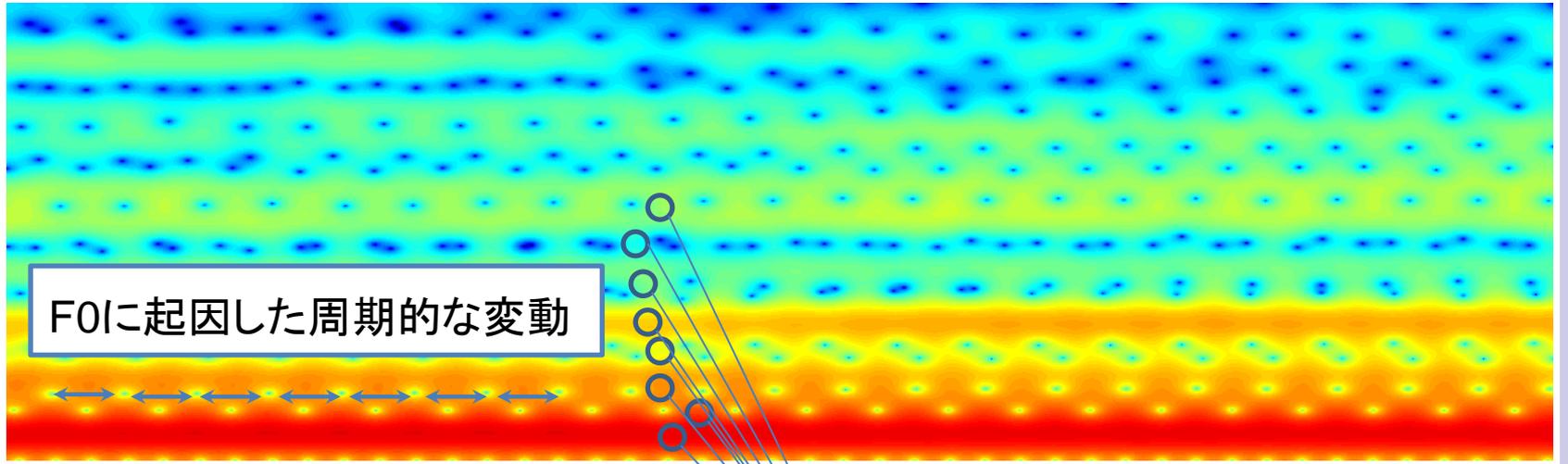
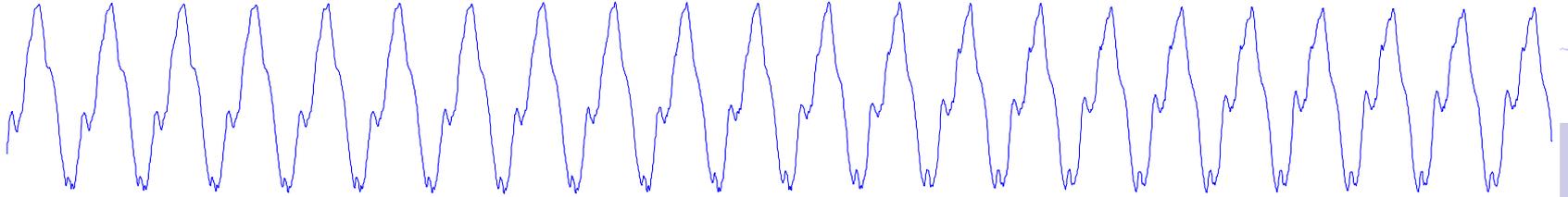
F0適応分析：F0適応スペクトル



F0適応分析：F0適応スペクトル(部分拡大)



F0適応スペクトルにも谷ができる



F0に起因した周期的な変動

周波数毎に共振時刻が異なる
= 群遅延
 (フォルマントの変動も関係)

提案手法：F0適応多重フレーム統合分析法

□ 特長

真の包絡の推定

ラベルを使わない

□ 二つの分析方法の組み合わせで実現

■ F0適応分析

真の包絡の推定

- 基本周期の2.0倍の長さの窓で分析（従来と同様）
- 全サンプリング点で分析する：速い変化を追従する

■ 多重フレーム統合分析

真の包絡の推定

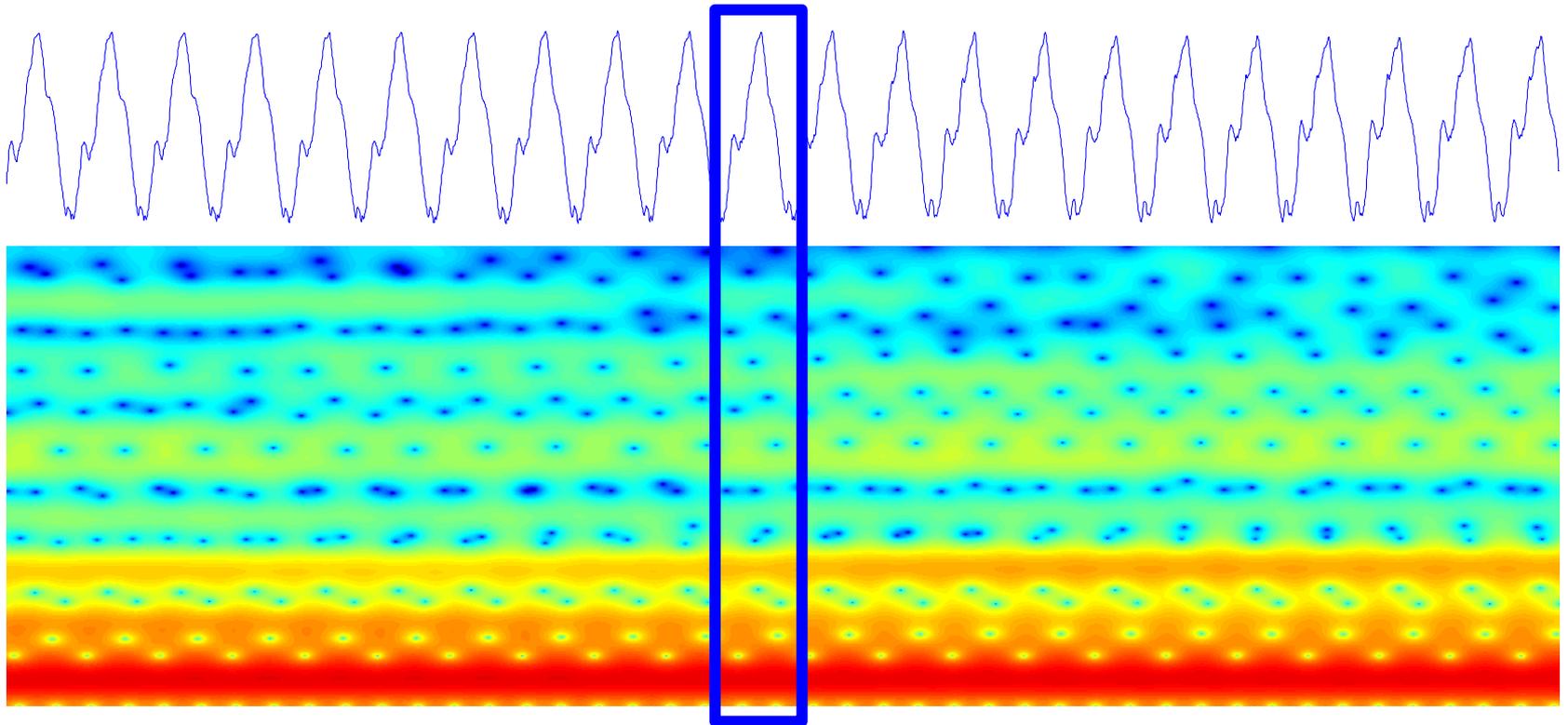
ラベルを使わない

- 近傍の複数フレームを統合する
 - 単一のピッチマークよりも頑健（スペクトルの谷が発生しない）
 - ピッチマークや音素ラベルが不要

多重フレーム統合分析

□ 基本周期の範囲で

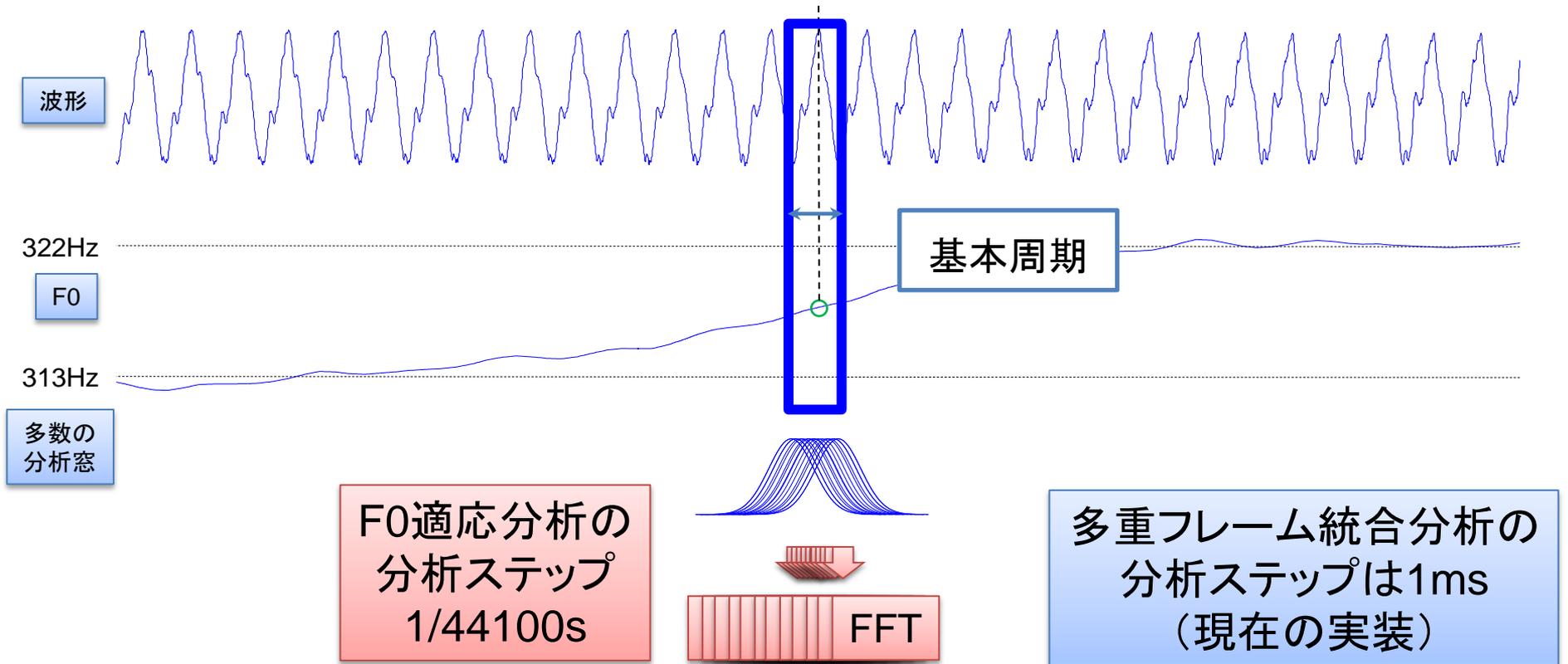
F0適応スペクトルを統合して変動を消去



多重フレーム統合分析:基本周期の範囲で統合

□ 基本周期の範囲で

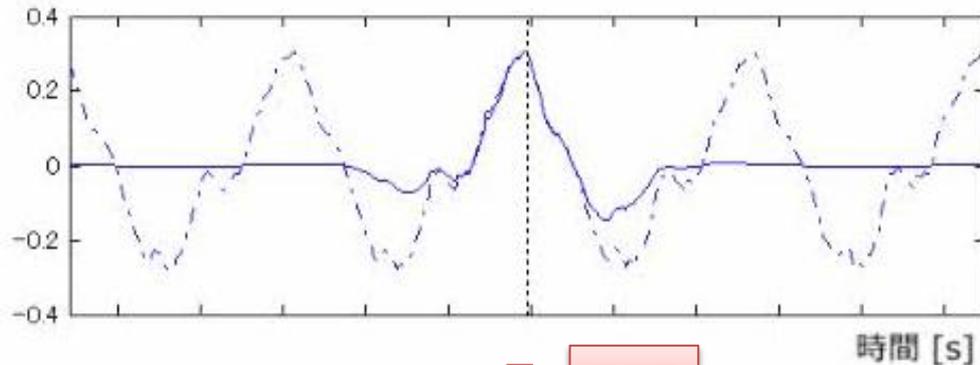
F0適応スペクトルを統合して変動を消去



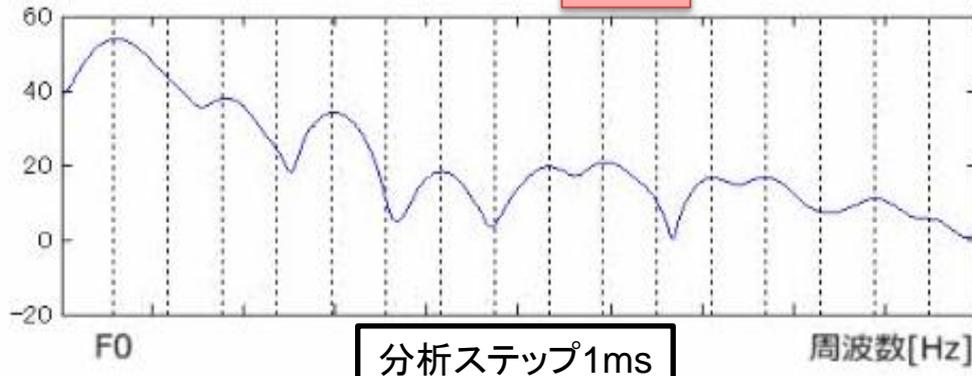
多重フレーム統合分析:基本周期の範囲で統合

□ 基本周期の範囲で

F0適応スペクトルを統合して変動を消去



↓ FFT



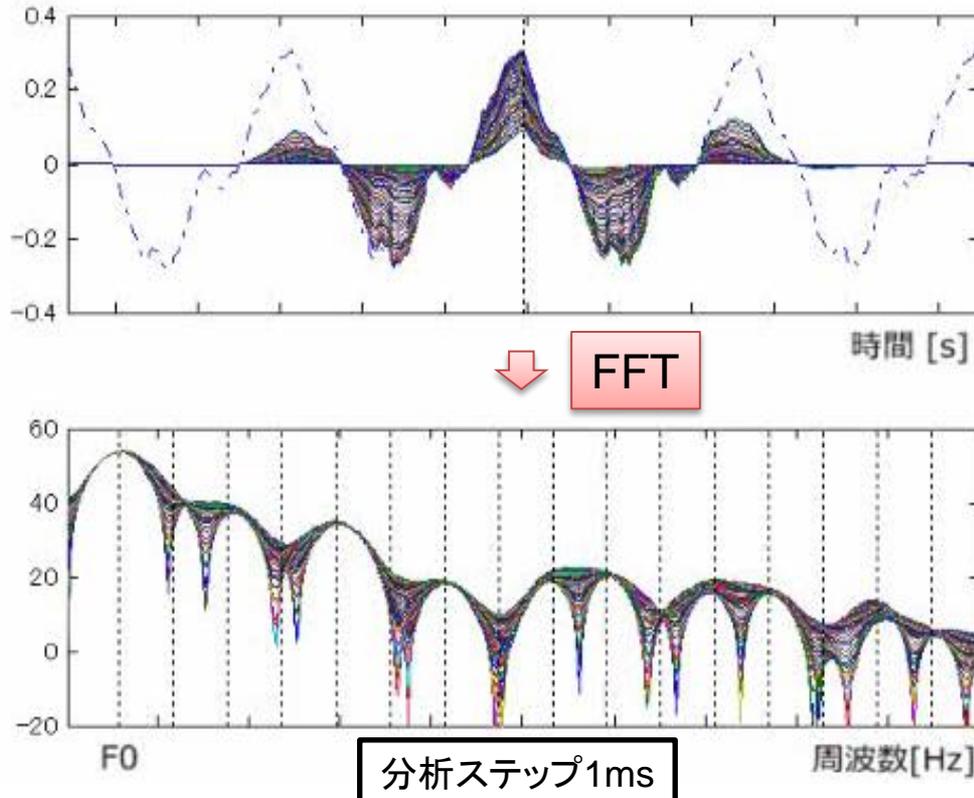
統合を
行わない場合

谷が発生

多重フレーム統合分析:基本周期の範囲で統合

□ 基本周期の範囲で

F0適応スペクトルを統合して変動を消去

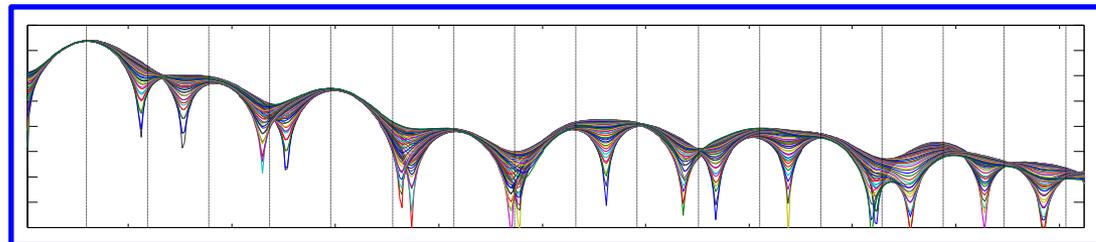


統合(重畳)
により

どこで分析しても
谷が埋まる

多重フレーム統合分析:基本周期の範囲で統合

- 統合:基本周期で重畳したF0適応スペクトル



から**スペクトル包絡**を推定

- 本提案での仮定

- 真のスペクトル包絡は

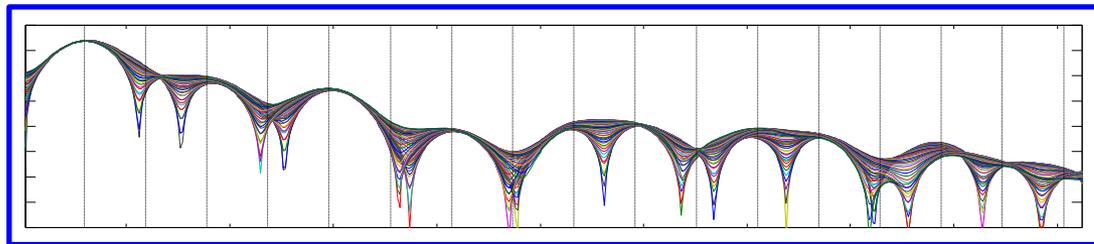
最大包絡と最小包絡の間に存在

- 最小包絡の**谷は観測できなかつた値**のため

推定時には利用しない

多重フレーム統合分析:基本周期の範囲で統合

- 統合:基本周期で重畳したF0適応スペクトル



から**スペクトル包絡**を推定

- 本提案での仮定

- 真のスペクトル包絡は

最大包絡と最小包絡の間に存在

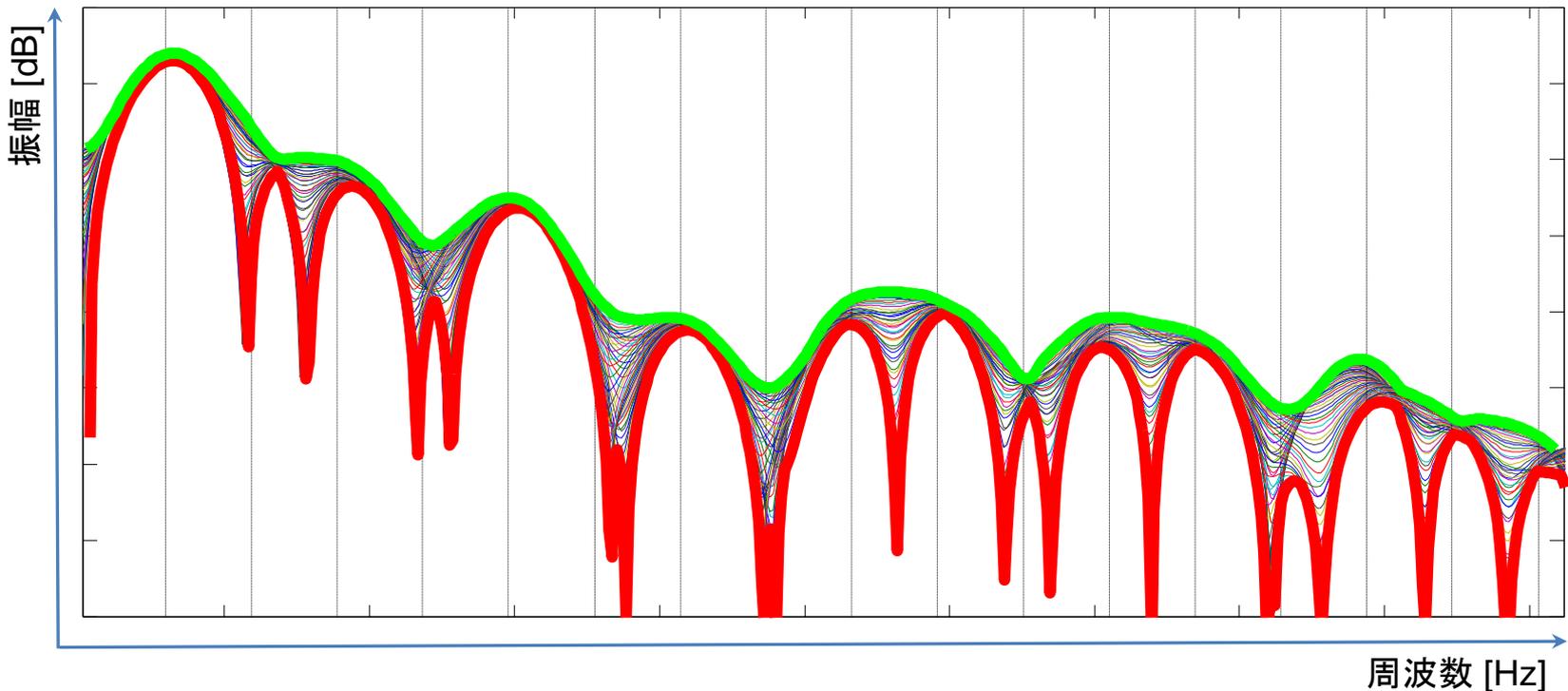
- 最小包絡の**谷**は観測できなかつた**値**のため

推定時には利用しない

多重フレーム統合分析:包絡線の存在範囲の設定

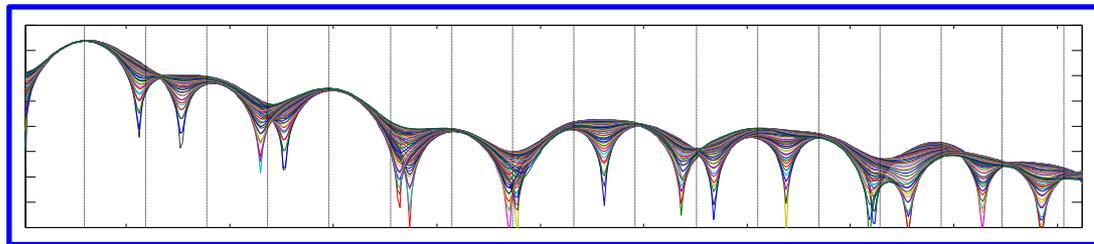
□ 基本周期で重畳したF0適応スペクトルの

最大包絡線 ~ 最小包絡線



多重フレーム統合分析:基本周期の範囲で統合

- 統合:基本周期で重畳したF0適応スペクトル



から**スペクトル包絡**を推定

- 本提案での仮定

- 真のスペクトル包絡は

最大包絡と最小包絡の間に存在

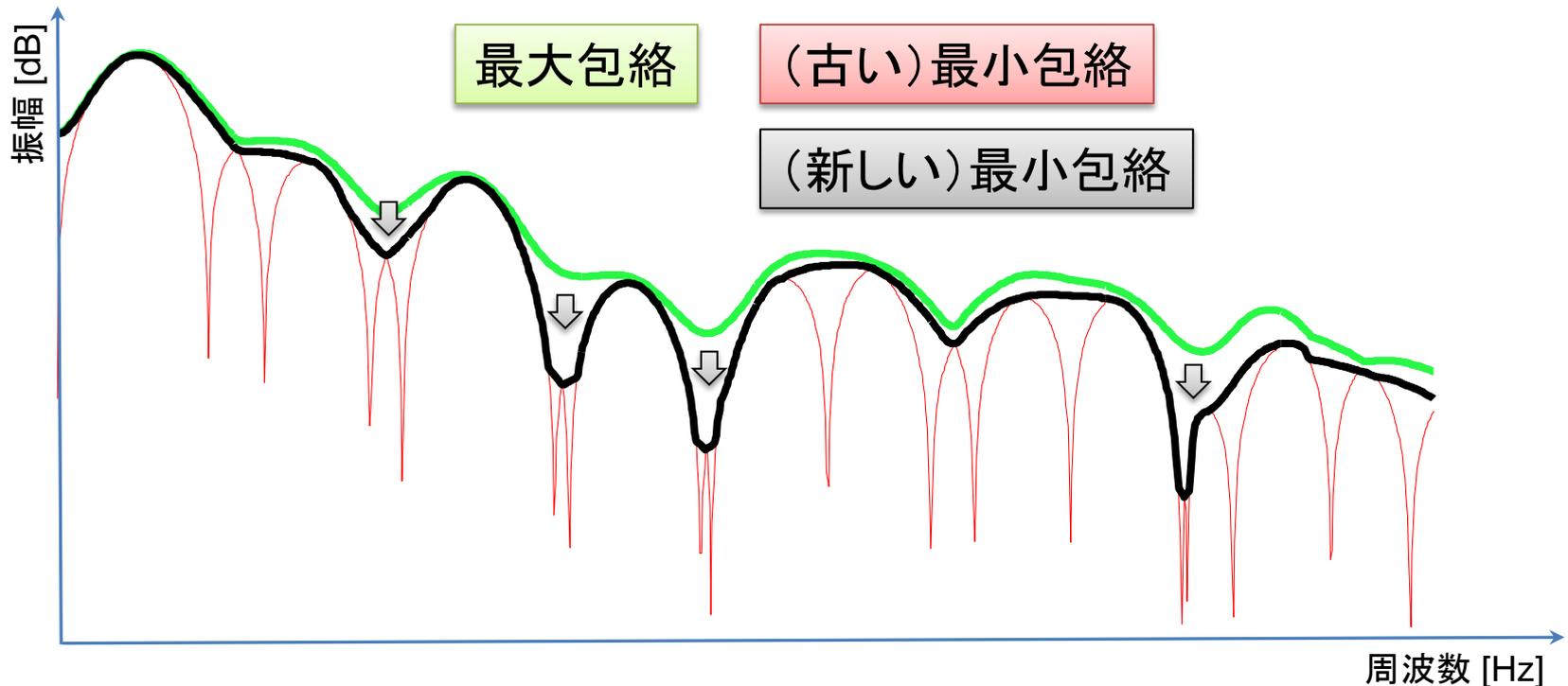
- 最小包絡の**谷は観測できなかつた値**のため

推定時には利用しない

多重フレーム統合分析:最小包絡の谷を除去

□ 谷を除去

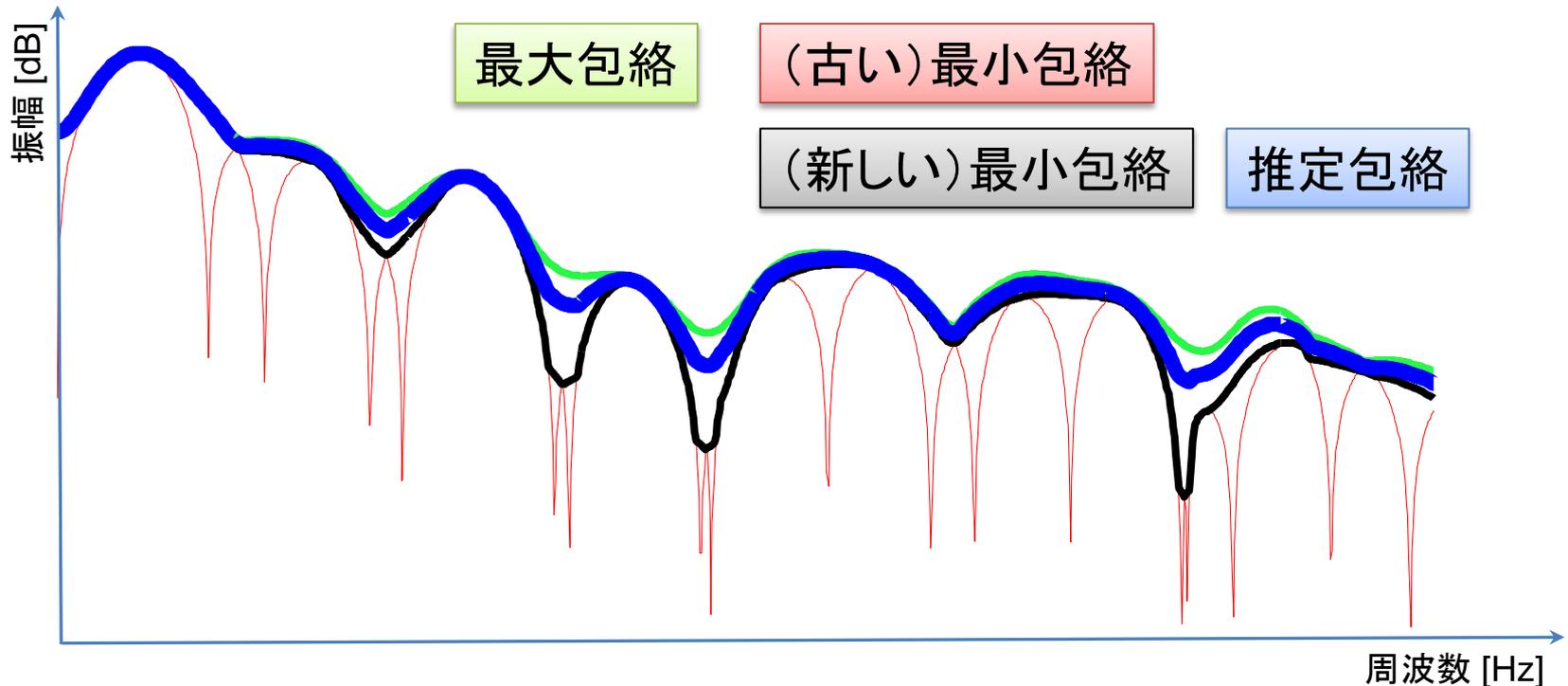
- **最大包絡を最小包絡にかぶせるように変形**
 - 除去後の最小包絡の形を、最大包絡に基づいて決定



多重フレーム統合分析:統合して包絡を推定

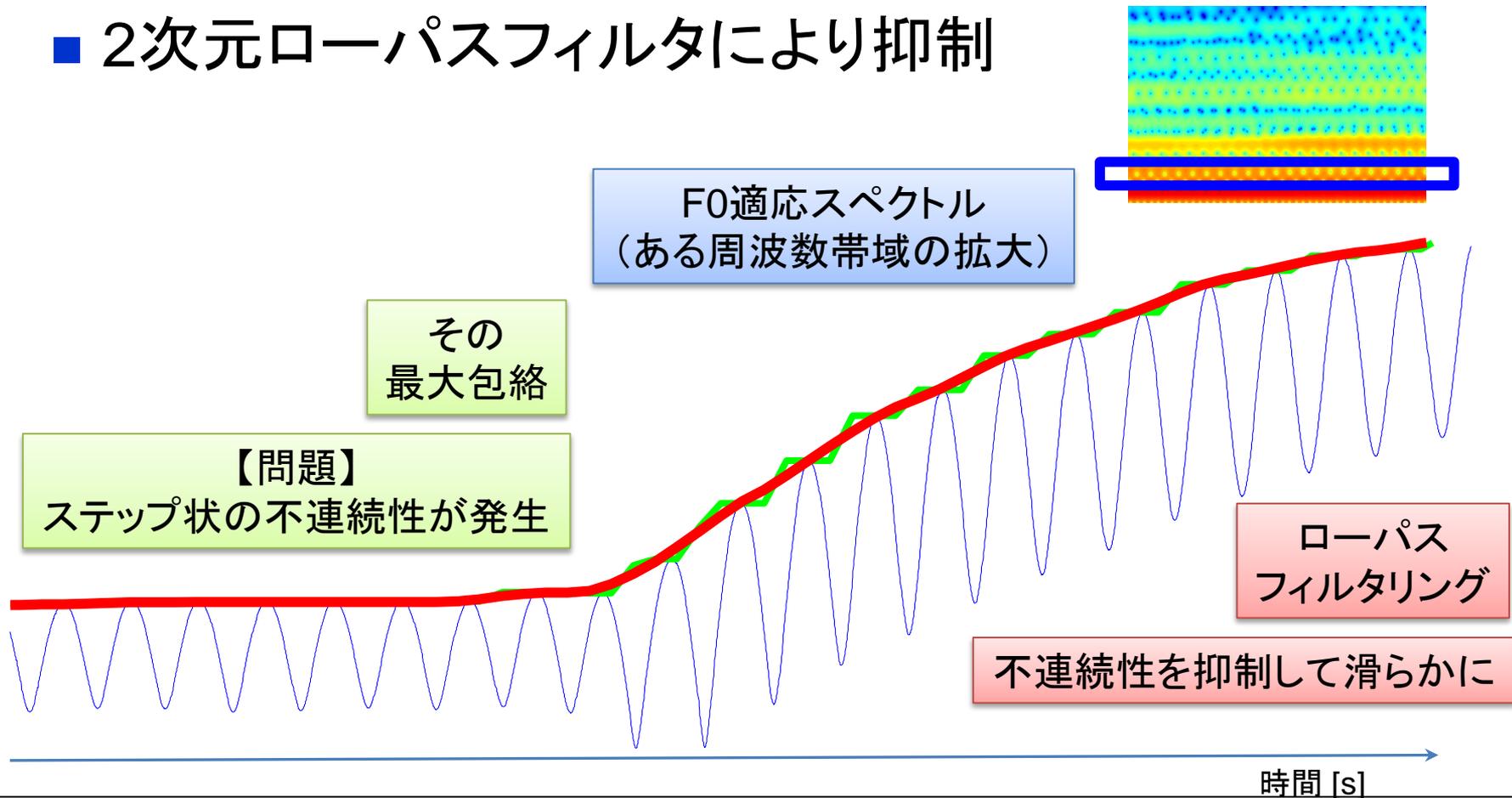
□ スペクトル包絡を推定

- 最大包絡と谷を除去した最小包絡との平均



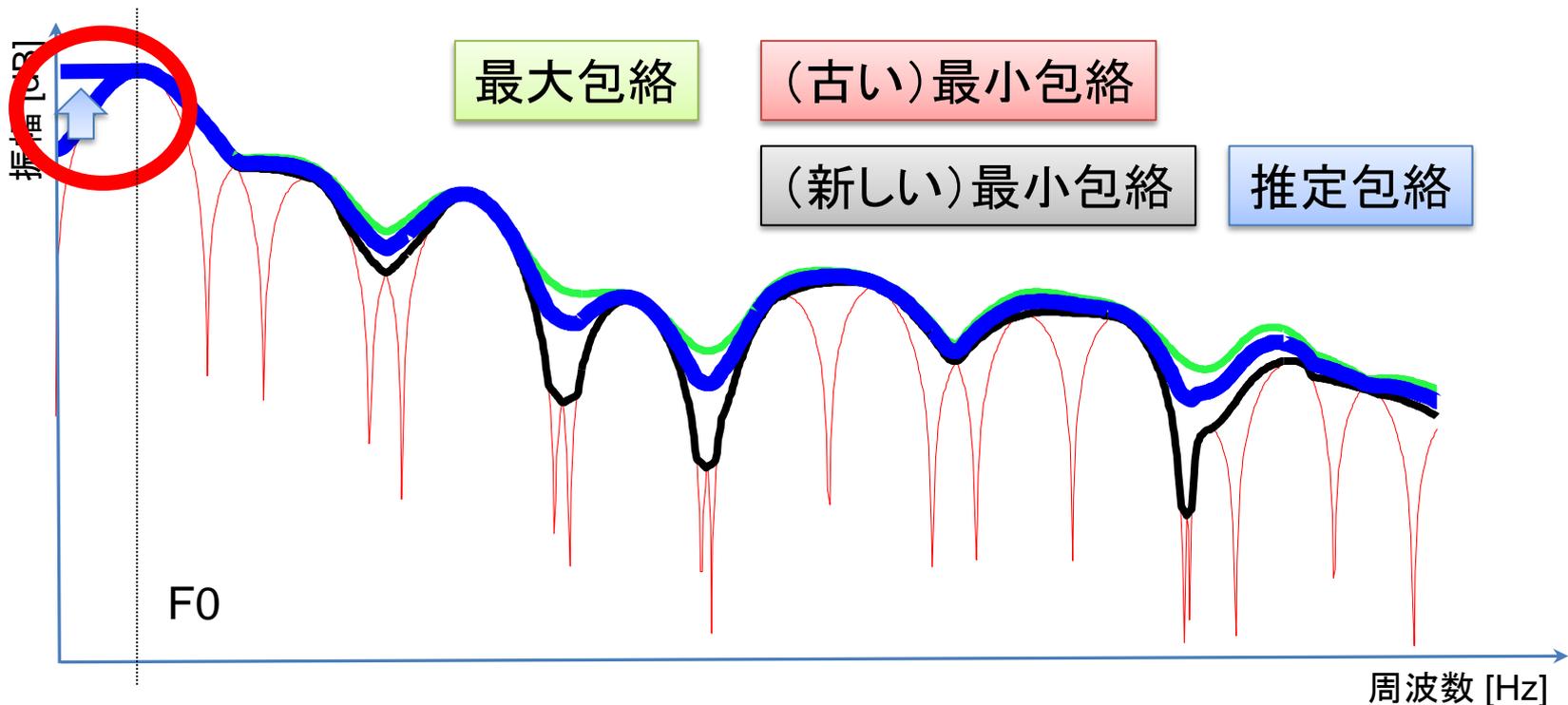
多重フレーム統合分析:精度向上のための処理(1)

- 最大操作によりステップ状の不連続性が発生
 - 2次元ローパスフィルタにより抑制



多重フレーム統合分析:精度向上のための処理(2)

- F0以下の包絡が安定して推定できない
 - F0以下の値はF0における値で置換



実験

実験条件(1)

□ 比較する従来手法

- **STRAIGHT, TANDEM-STRAIGHT**
[Ver. V40_007d] [Ver. V009x]

□ 対象とする音

■ STRAIGHTで分析再合成した音

- 男性の**歌声** (ソロ) : RWC 研究用音楽DB, G, N0.91 [後藤 他, 2004]
- 女性の**話声** : AISTハミングDB, E008 [後藤 他, 2005]
- **楽器音** : ピアノ, バイオリン : RWC 研究用音楽DB, I, No.1&16 [後藤 他, 2004]
[Klatt, 1980]

■ Klatt合成器で包絡を生成して正弦波合成した音

- F1: 250, 250, 500, 1000, 1000, 500 Hz
- F2: 750, 1500, 1500, 1500, 2000, 2000 Hz

実験条件(2)

□ 提案手法の分析窓

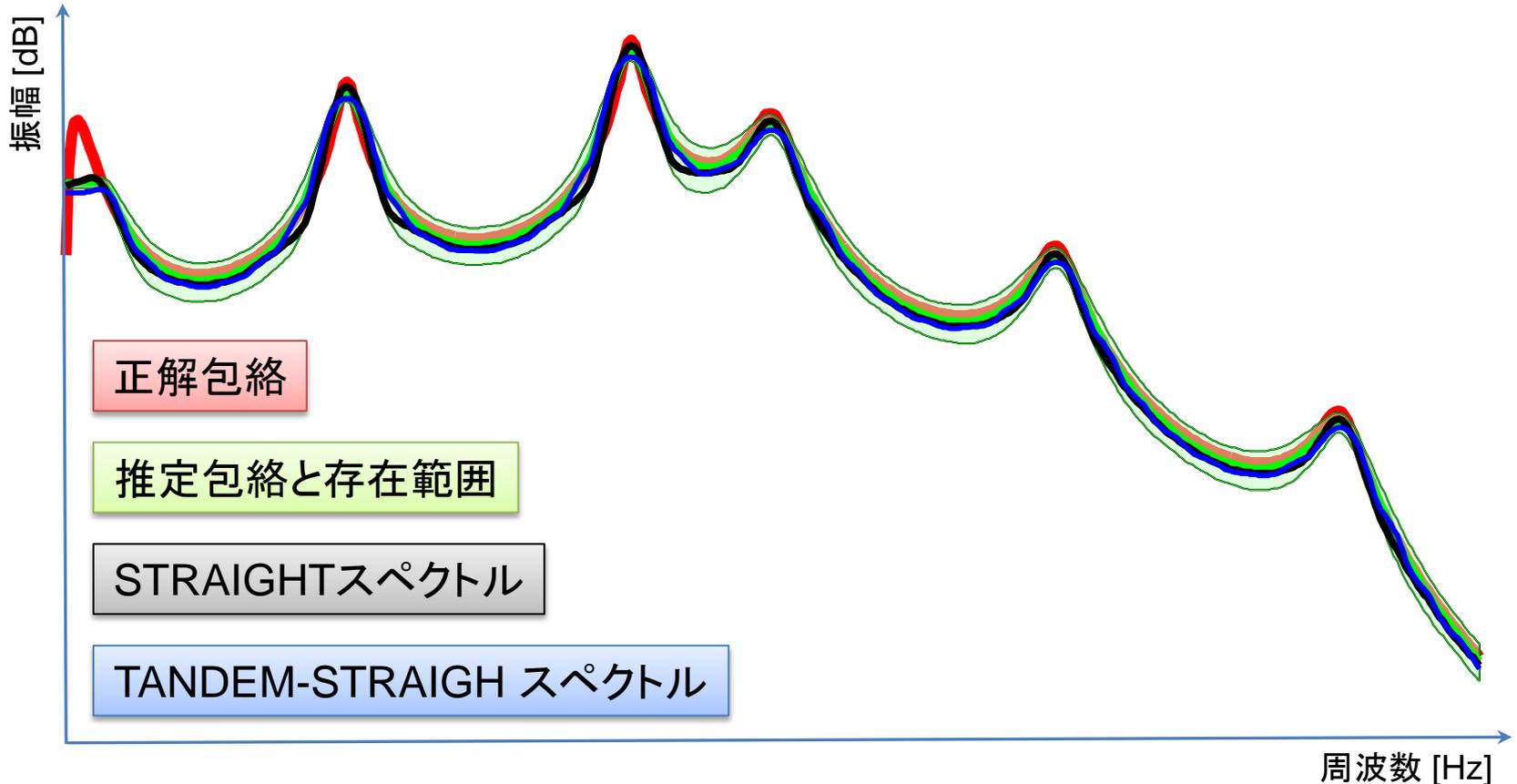
- 窓: F0適応させた**ガウス窓** (4096点FFT)
 - $3\sigma=1/F0$
- **F0適応分析**のステップ: 1/44100 s
- **多重フレーム統合分析**のステップ: 1 ms

□ 距離尺度

- 正解包絡との**対数スペクトル距離**
 - **フレーム毎に距離(線形振幅)を最小化してから積算した**
※形状を評価するため

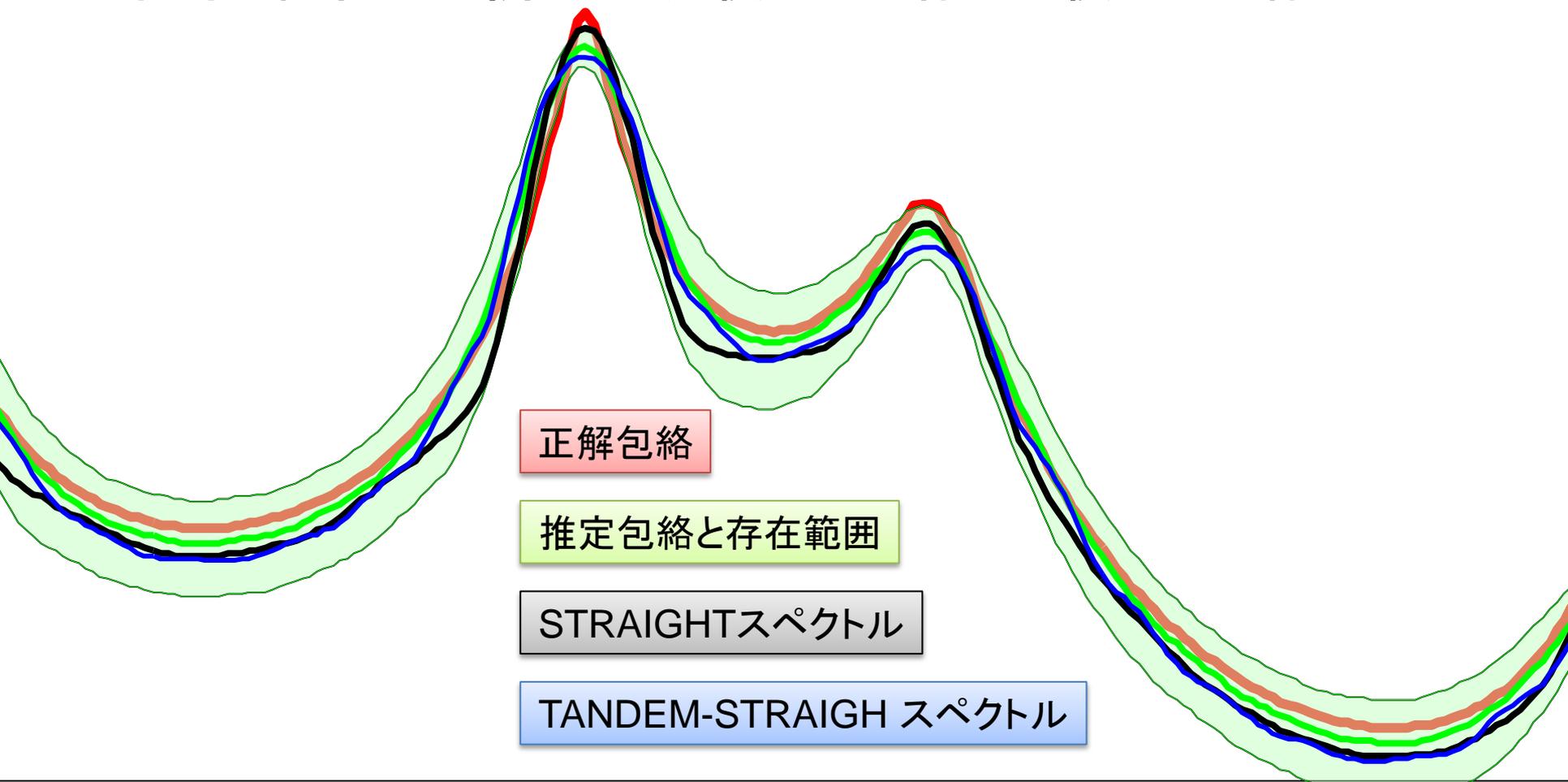
スペクトル包絡と存在範囲の推定結果

- 存在範囲：(新しい)最小包絡～最大包絡

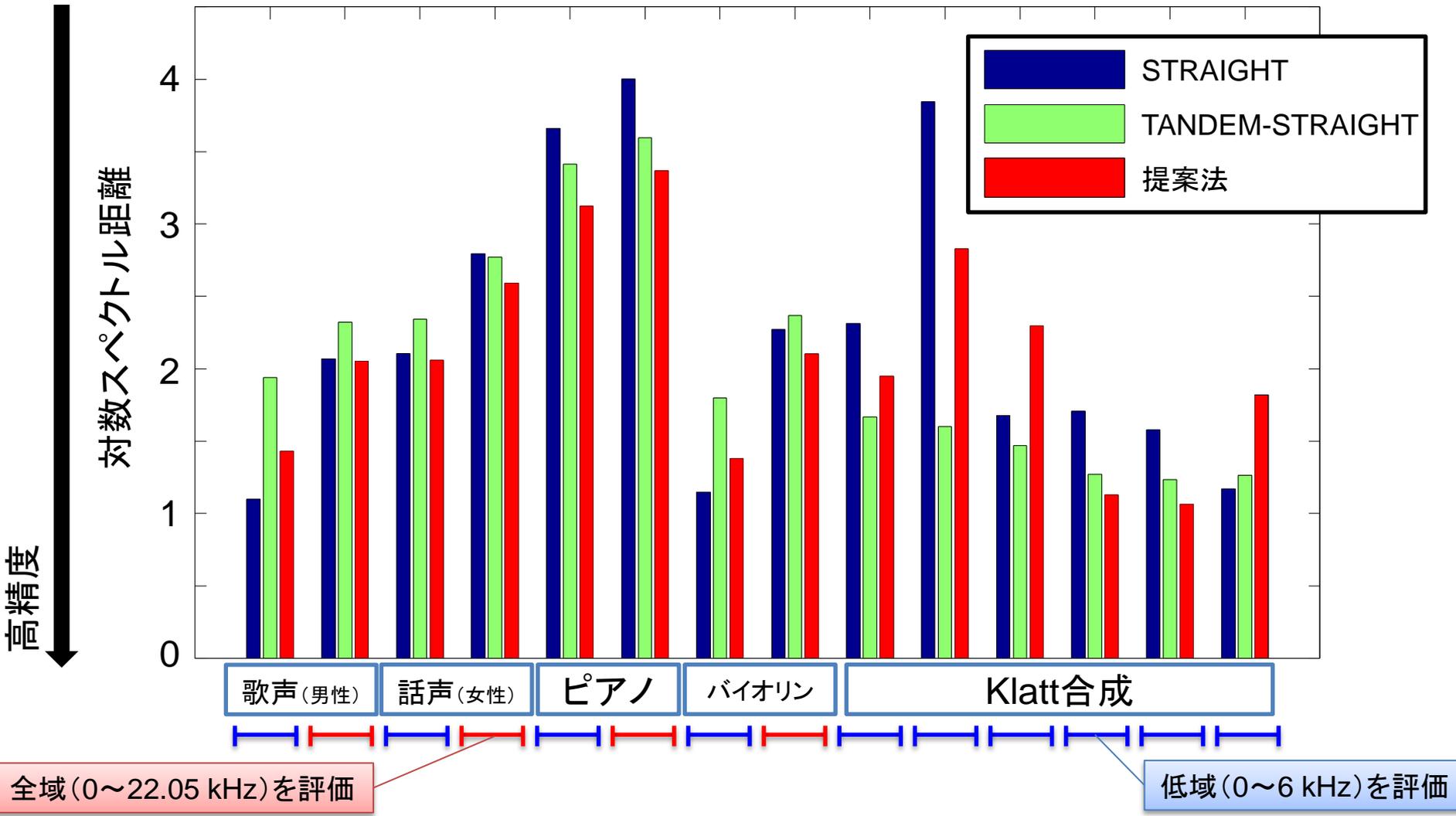


スペクトル包絡と存在範囲の推定結果 (拡大)

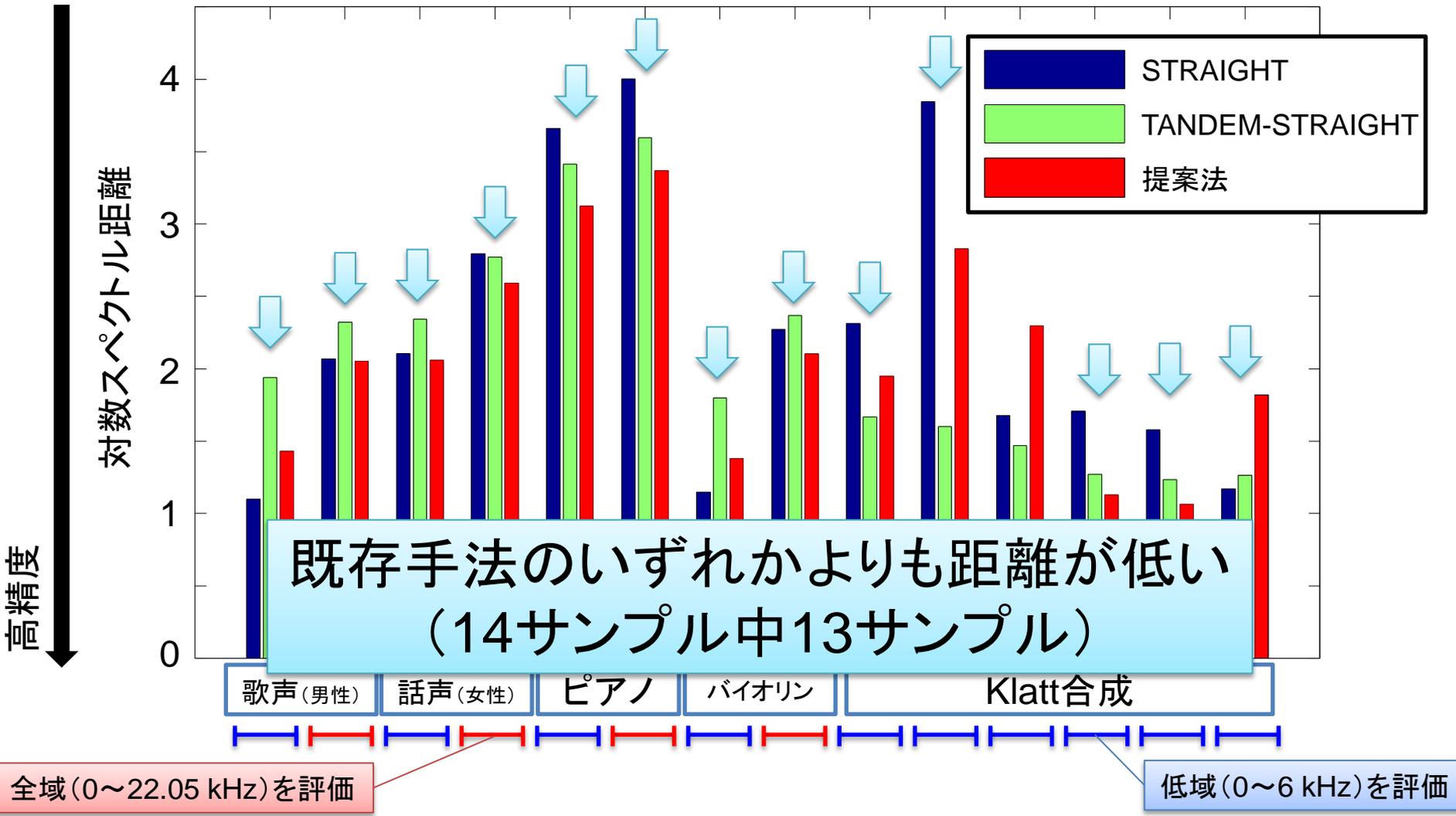
- 存在範囲: (新しい) 最小包絡 ~ 最大包絡



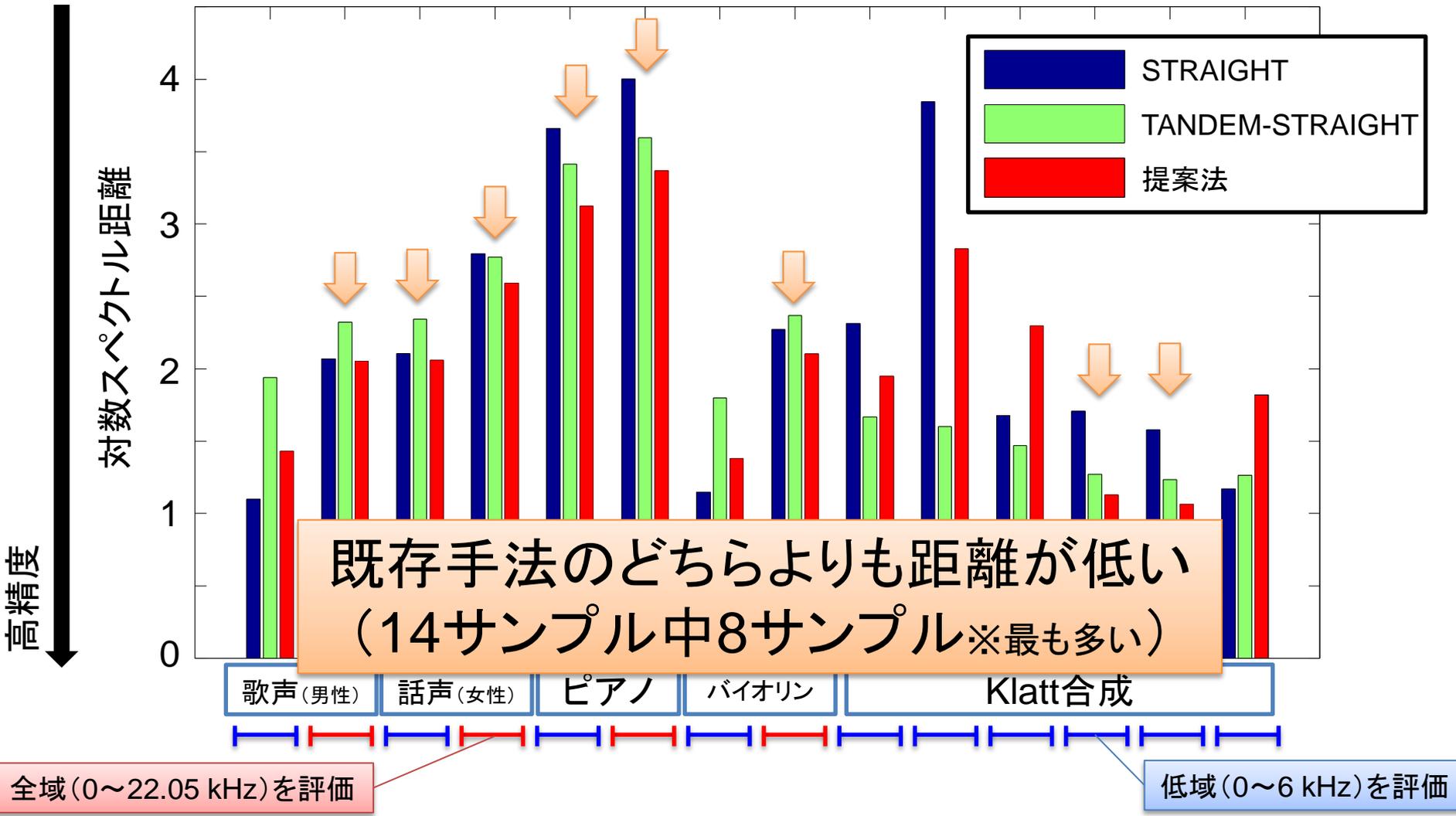
スペクトル包絡の推定精度



スペクトル包絡の推定精度



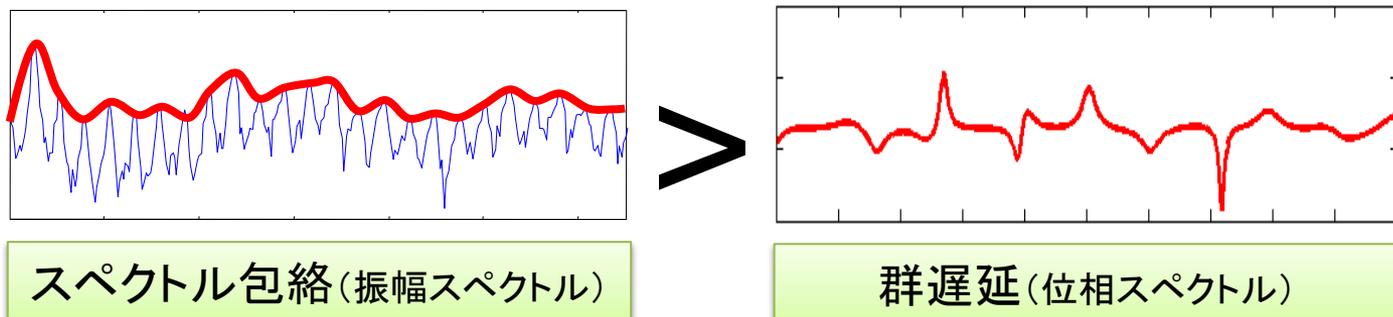
スペクトル包絡の推定精度



F0適応多重フレーム統合分析に基づく 群遅延(位相情報)の推定

群遅延の必要性

□ 音の分析では振幅スペクトルが重視



□ 音の合成における自然知覚には位相も重要！

■ 正弦波合成:

[伊藤 他, 2008]

- 初期位相が $\pi/8$ 以上ずれると自然性が単調減少

■ 分析合成:

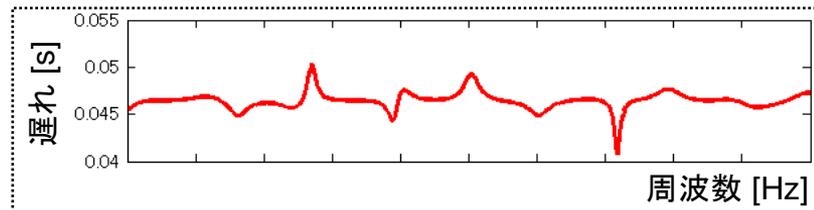
[松原 他, 2010]

- 単位波形の生成で最小位相応答が零位相応答より自然

群遅延とは

□ 位相スペクトルに関する情報 (周波数微分、符号反転)

- 周波数帯域毎の遅れ時間
 - 単位は秒
 - 各周波数での時間方向の重心



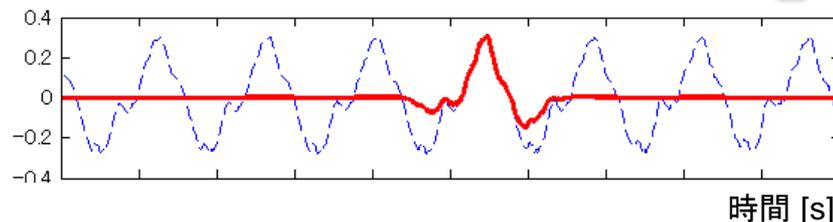
- 音源特性 (単位波形の形状)・音の質感を反映

□ 計算方法 (従来)

- **単位波形** に対して分析する

ピッチマークが必要

- FFTによる直接計算が可能 [坂野 他, 2000]
 - 位相接続 (phase unwrapping) が不要

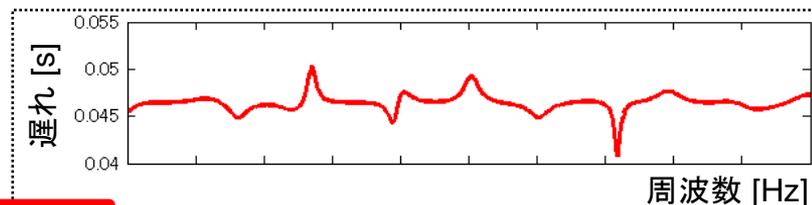


群遅延とは

□ 位相スペクトルに関する情報 (周波数微分、符号反転)

■ 周波数帯域毎の遅れ時間

- 単位は秒
- 各周波数での時間方向の重心

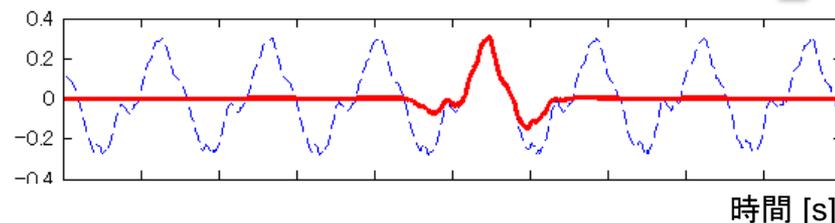


■ 音源特性 (単位波形の形状)・音の質感を反映

□ 計算方法 (従来)

■ 単位波形に対して分析する

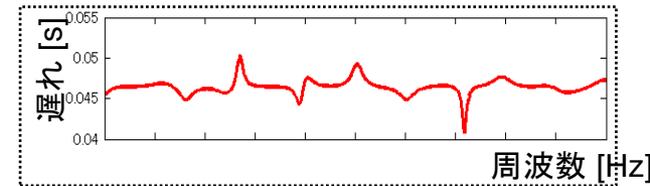
- FFTによる直接計算が可能 [坂野 他, 2000]
 - 位相接続 (phase unwrapping) が不要



ピッチマークが必要

群遅延とは

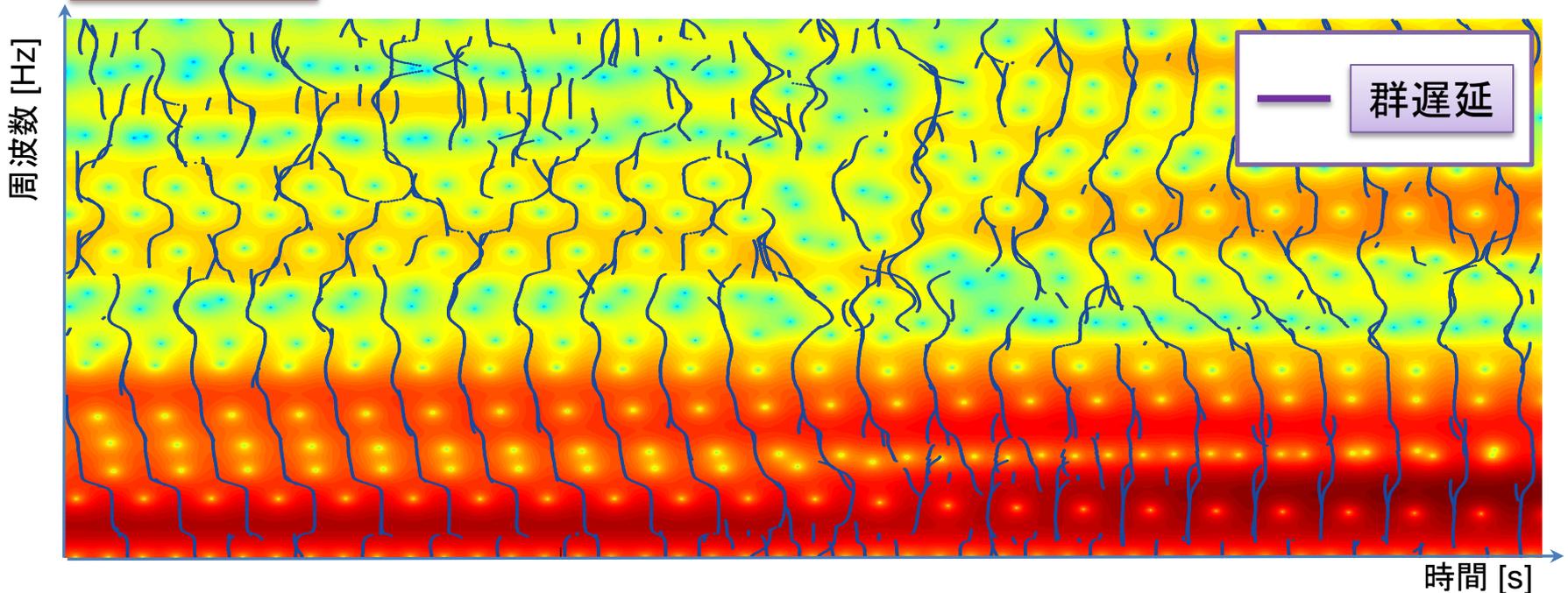
□ 各周波数での時間方向の重心



歌声波形



F0適応スペクトル

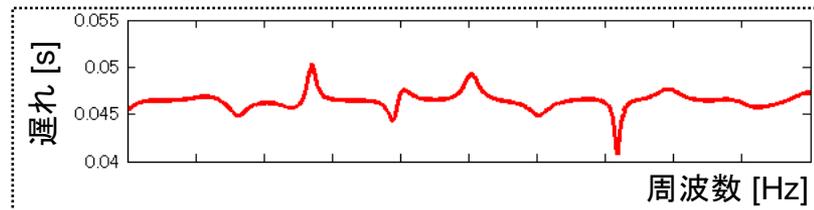


群遅延とは

□ 位相スペクトルに関する情報 (周波数微分、符号反転)

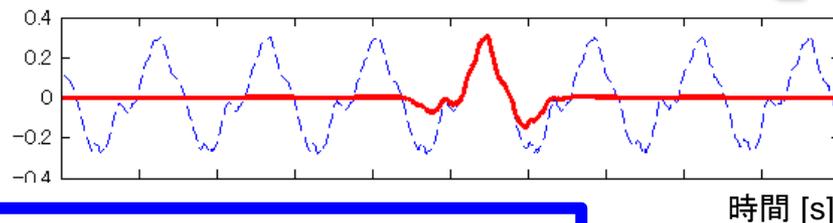
■ 周波数帯域毎の遅れ時間

- 単位は秒
- 各周波数での時間方向の重心



■ 音源特性 (単位波形の形状)・音の質感を反映

□ 計算方法 (従来)



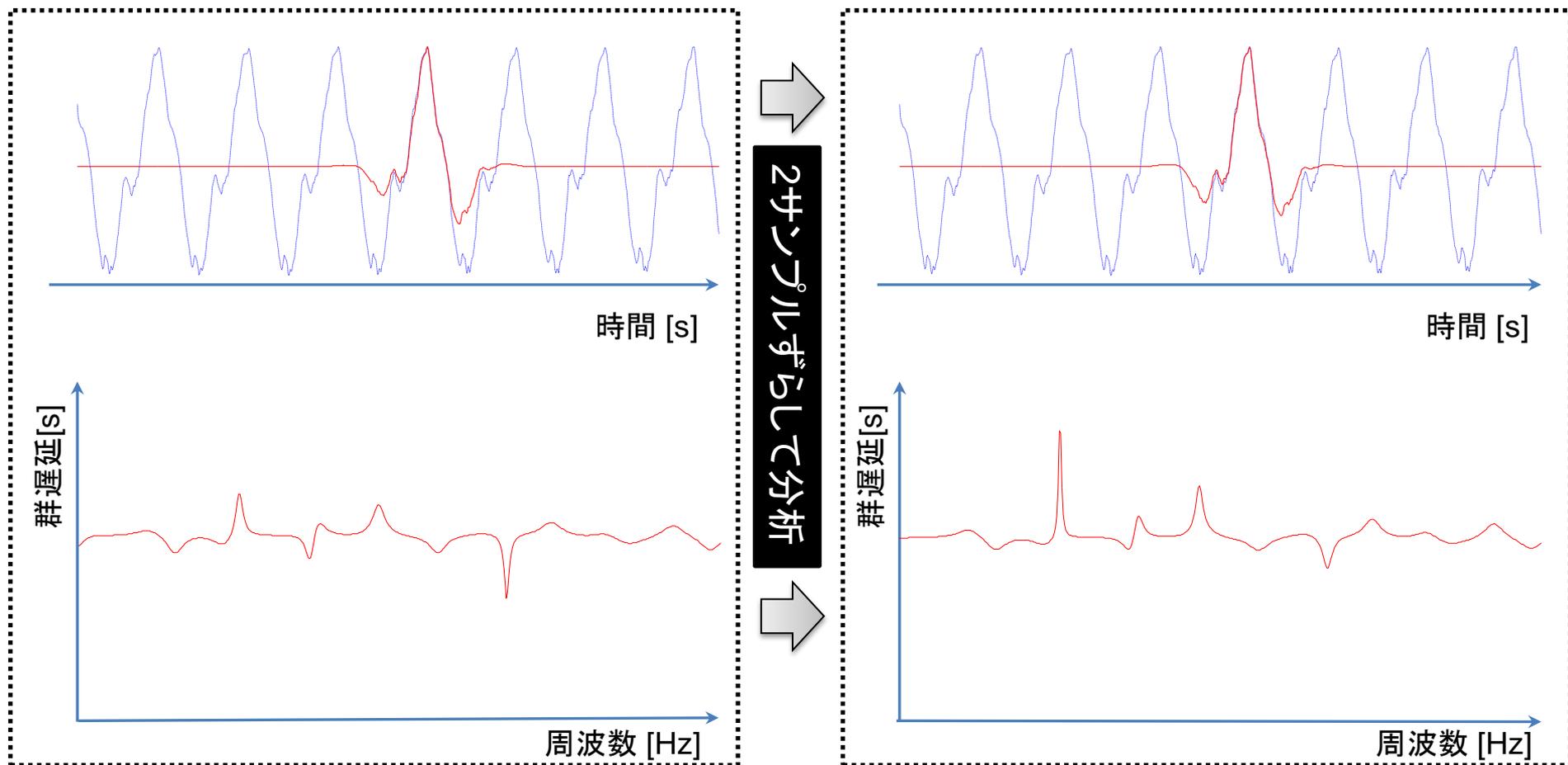
■ **単位波形** に対して分析する

ピッチマークが必要

- FFTによる直接計算が可能 [坂野 他, 2000]
 - 位相接続 (phase unwrapping) が不要

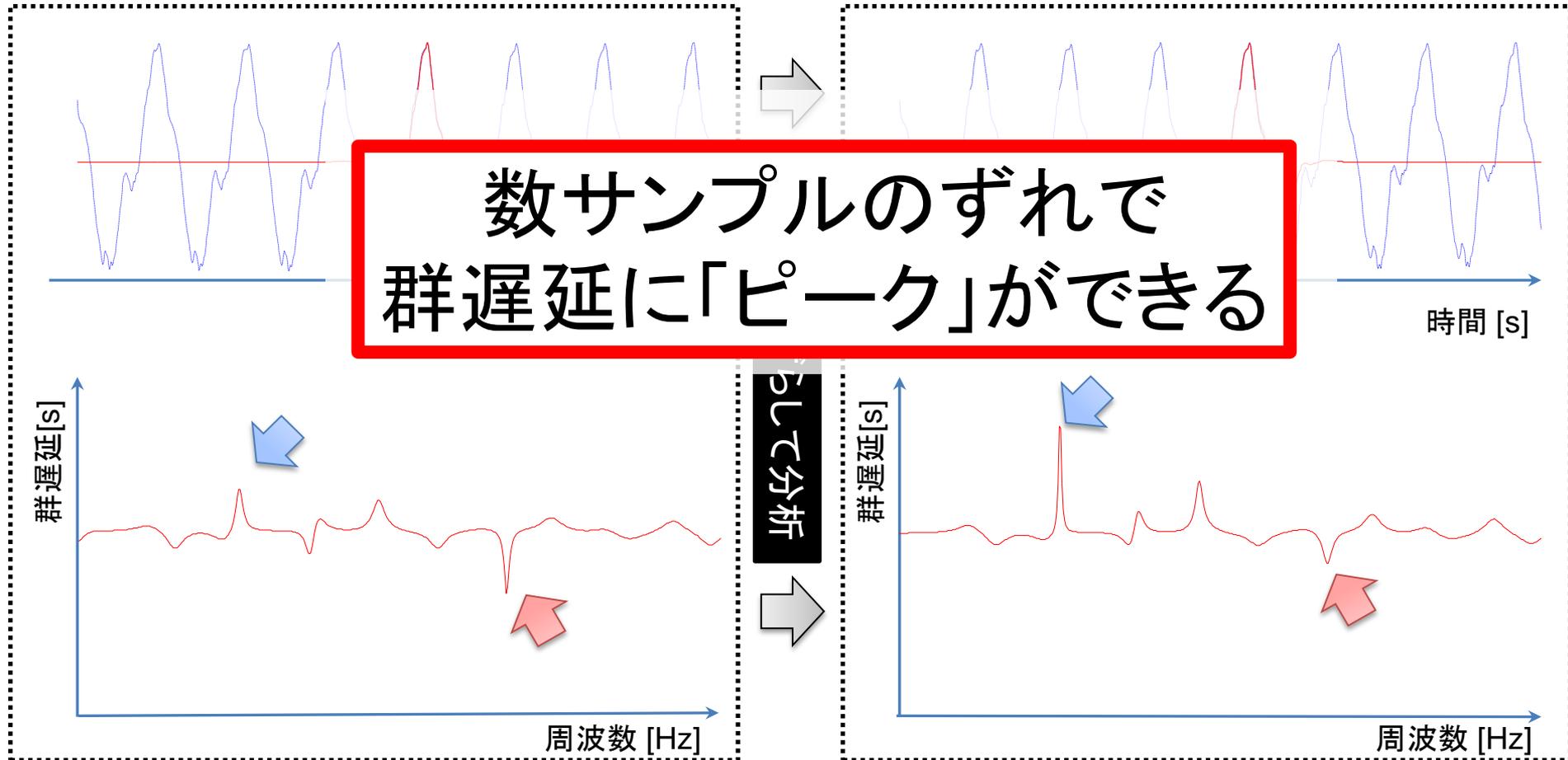
ピッチマーク分析の問題点(群遅延)

- 単一のピッチマークから真の値が得られない



ピッチマーク分析の問題点（群遅延）

- 単一のピッチマークから真の値が得られない



ピッチマーク分析の問題点（群遅延）

- 単一のピッチマークから真の値が得られない

これらピークは平滑化によって除去しても
品質に影響がない [坂野 他, 2000]

これらピークは
適切に観測できていないことが原因
と考えられる

包絡推定と位相の合成への応用(従来)

[坂野 他, ,2000,坂野 他, ,2001]

□ 単位波形から群遅延を推定して平滑化

- 平滑化によるピークの除去と次元削減が可能
- ピッチマークが必要



□ PLATINUM [Morise, ,2012]

- スペクトル包絡の最小位相応答と
元の単位波形との逆畳込みによる励振源の抽出
- ピッチマークが必要
- 表現方法が直接的(波形そのもの=モデル化が難しい)

その他: PSOLA, 正弦波モデル

提案手法: F0適応多重フレーム統合分析法

□ 特長

真の群遅延の推定

ラベルを使わない

□ 二つの分析方法の組み合わせで実現

■ F0適応分析

真の群遅延の推定

- 基本周期の2.0倍の長さの窓で分析(従来と同様)
- 全サンプリング点で分析する: 速い変化を追従する

■ 多重フレーム統合分析

真の群遅延の推定

ラベルを使わない

- 近傍の複数フレームを統合する
 - 単一のピッチマークよりも頑健 (群遅延のピークが発生しない)
 - ピッチマークや音素ラベルが不要

提案手法：F0適応多重フレーム統合分析法

□ 特長 真の群遅延の推定 ラベルを使わない

□ 二つの分析方法の組み合わせで実現

■ F0適応分析 真の群遅延の推定

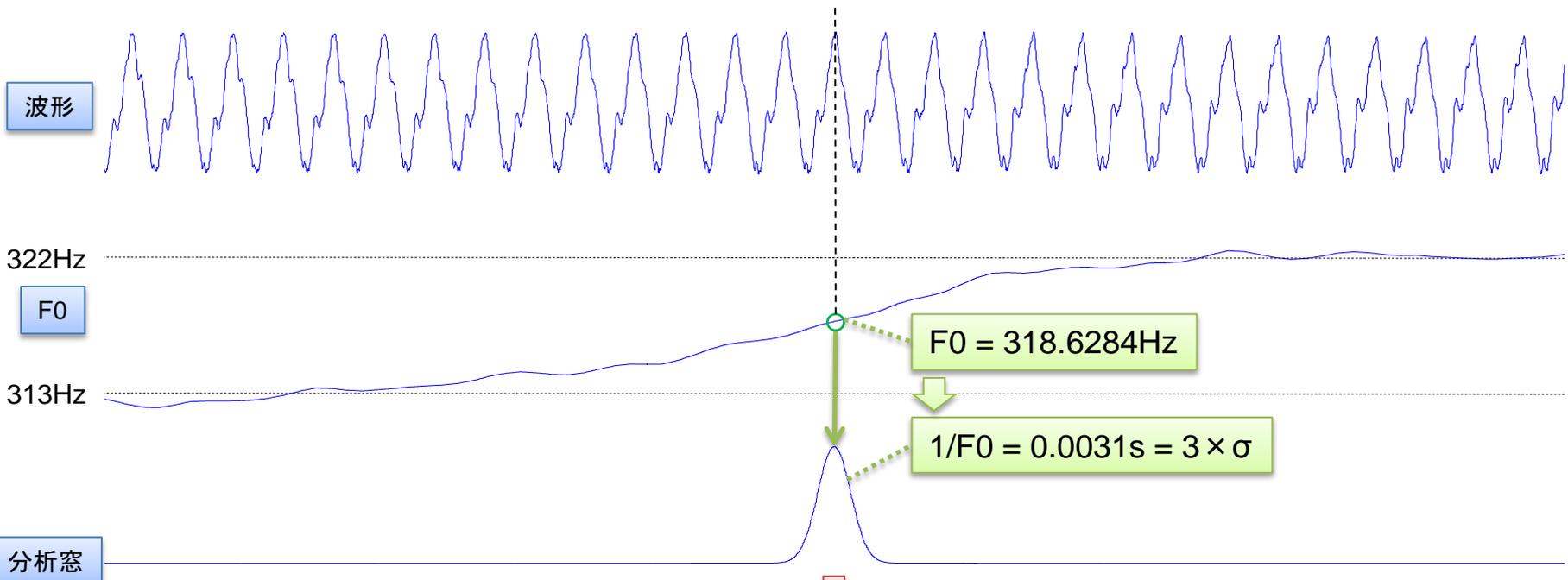
- 基本周期の2.0倍の長さの窓で分析（従来と同様）
- 全サンプリング点で分析する：速い変化を追従する

■ 多重フレーム統合分析 真の群遅延の推定 ラベルを使わない

- 近傍の複数フレームを統合する
 - 単一のピッチマークよりも頑健 群遅延のピークが発生しない
 - ピッチマークや音素ラベルが不要

F0適応分析:F0に適応した長さの短い窓

- F0に適応した長さの短い窓で
全時間(全サンプリング点)に対してFFT



群遅延抽出アルゴリズム

提案手法: F0適応多重フレーム統合分析法

□ 特長 真の群遅延の推定 ラベルを使わない

□ 二つの分析方法の組み合わせで実現

■ F0適応分析 真の群遅延の推定

- 基本周期の2.0倍の長さの窓で分析(従来と同様)
- 全サンプリング点で分析する: 速い変化を追従する

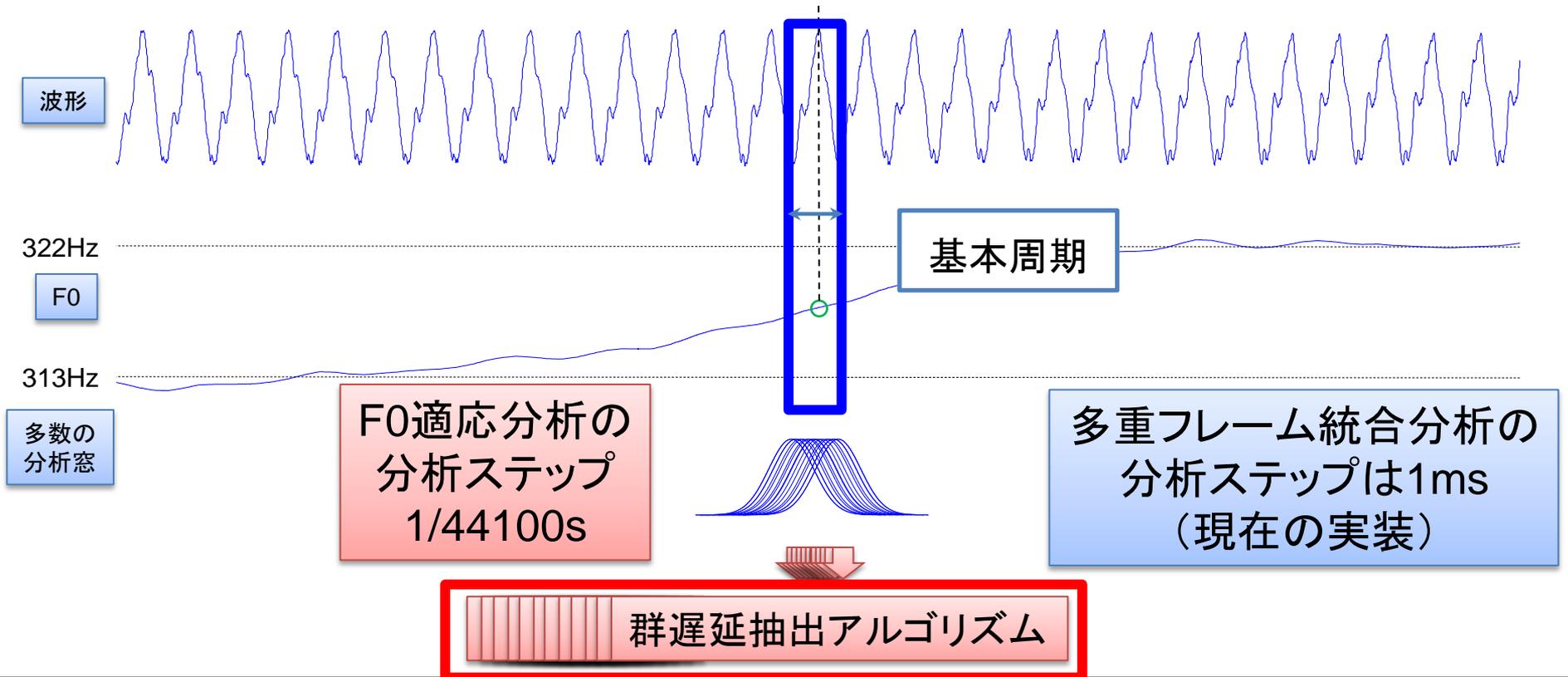
■ 多重フレーム統合分析 真の群遅延の推定 ラベルを使わない

- 近傍の複数フレームを統合する
 - 単一のピッチマークよりも頑健 (群遅延のピークが発生しない)
 - ピッチマークや音素ラベルが不要

多重フレーム統合分析:基本周期の範囲で統合

□ 基本周期の範囲で

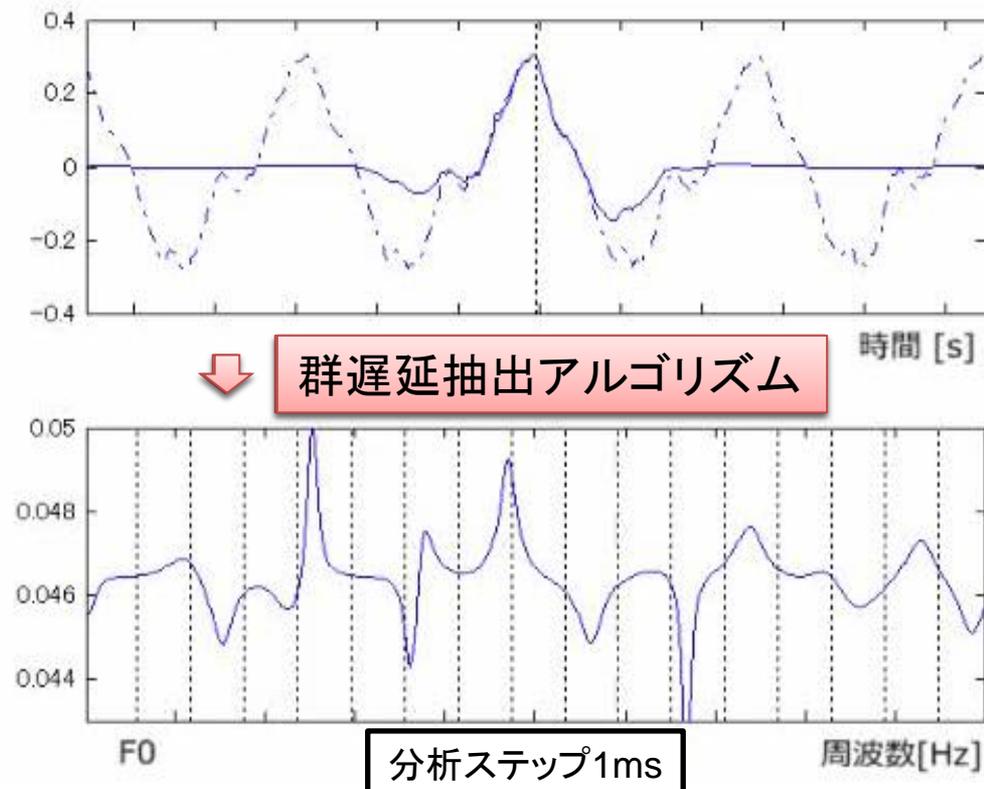
F0適応群遅延を統合して変動を消去



多重フレーム統合分析:基本周期の範囲で統合

□ 基本周期の範囲で

F0適応群遅延を統合して変動を消去



統合を
行わない場合

ピークが発生

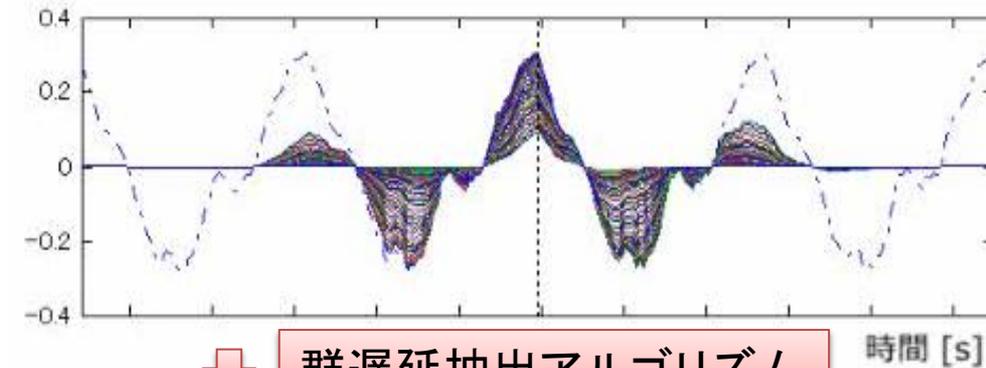
多重フレーム統合分析:基本周期の範囲で統合

□ 基本周期の範囲で

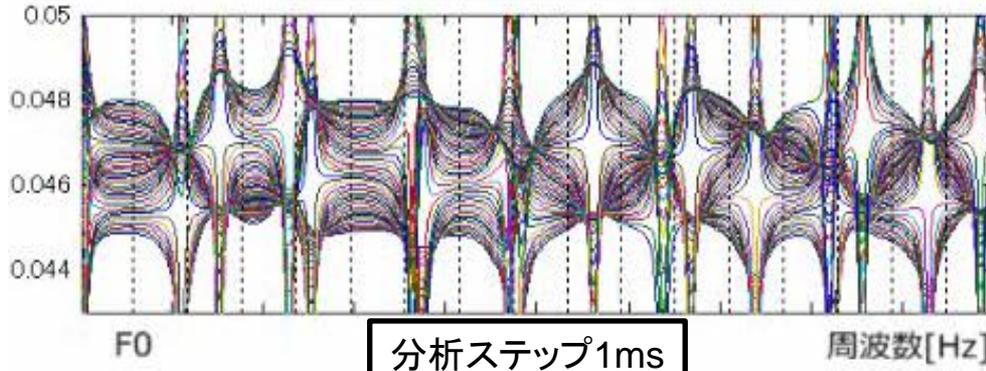
F0適応群遅延を統合して変動を消去

統合(重畳)
する

どこを選ぶか？

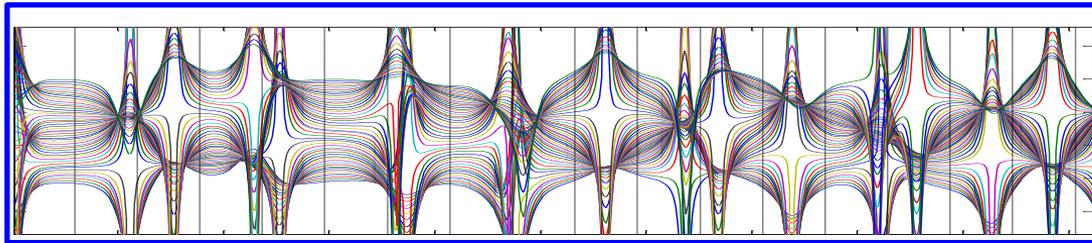


群遅延抽出アルゴリズム



多重フレーム統合分析:基本周期の範囲で統合

- 統合:基本周期で重畳したF0適応群遅延



から安定した群遅延を推定

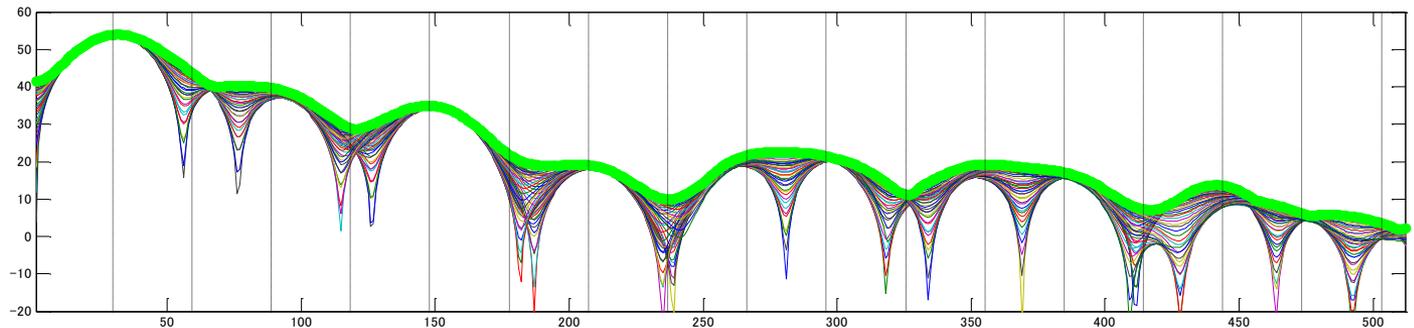
- 本提案での仮定
 - 真の群遅延はスペクトルの最大包絡に対応
 - (理由)最大包絡は時間方向の重心を表現するため

多重フレーム統合分析:基本周期の範囲で統合

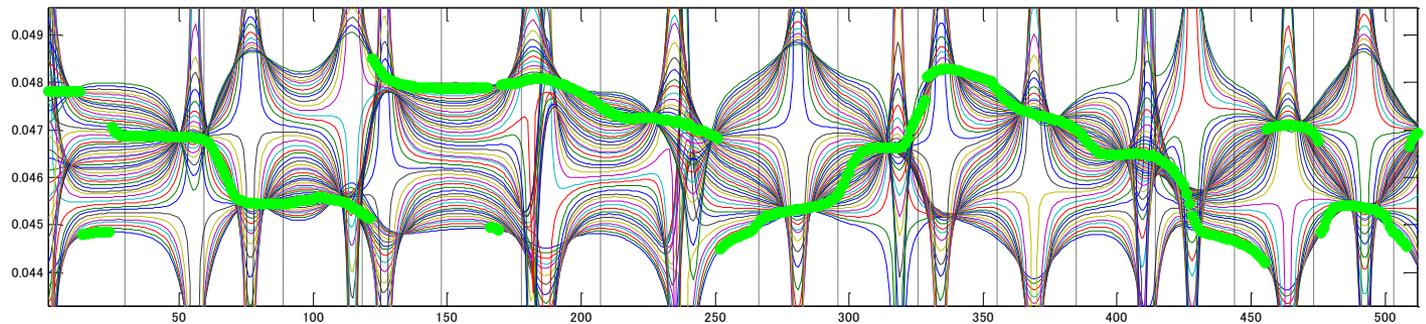
□ 基本周期の範囲で

最大包絡に対応する群遅延を選択

F0適応スペクトル
と
最大包絡



最大包絡
に対応する
群遅延



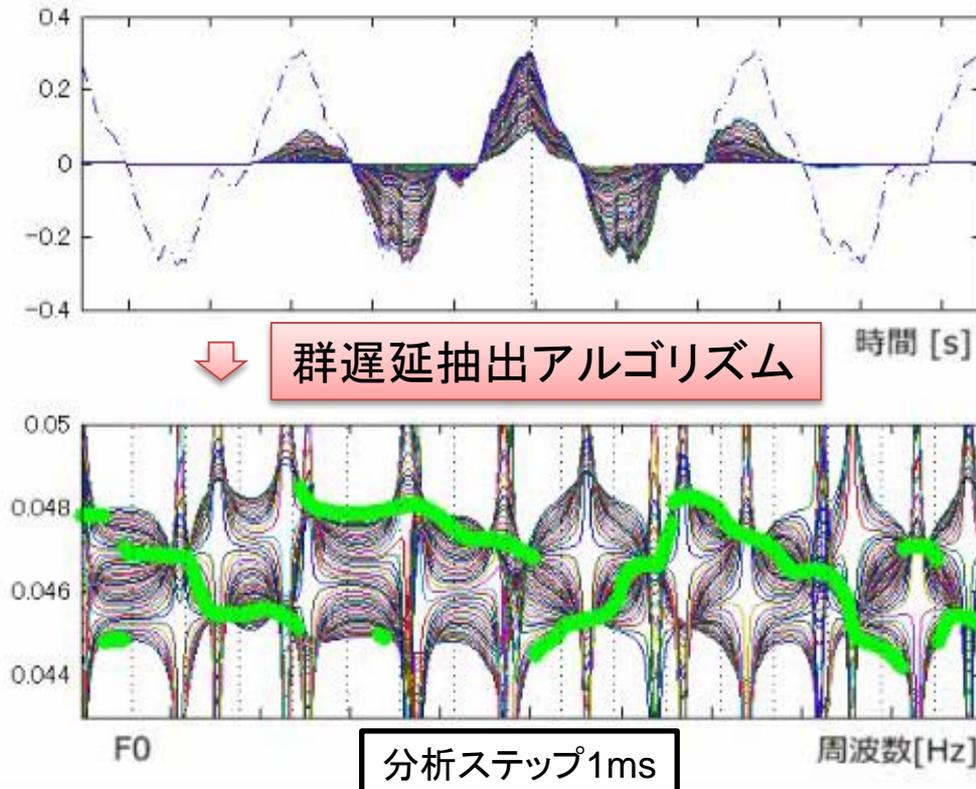
多重フレーム統合分析:基本周期の範囲で統合

□ 基本周期の範囲で

F0適応群遅延を統合して変動を消去

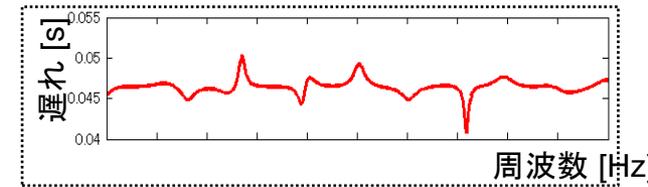
統合(重畳)
により

どこで分析しても
相対的に同じ形状

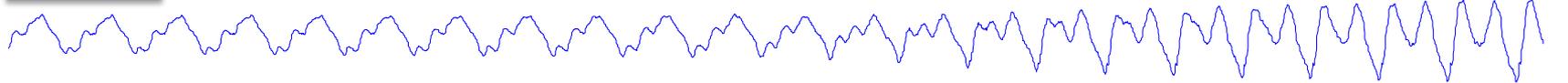


(再掲) 群遅延とは

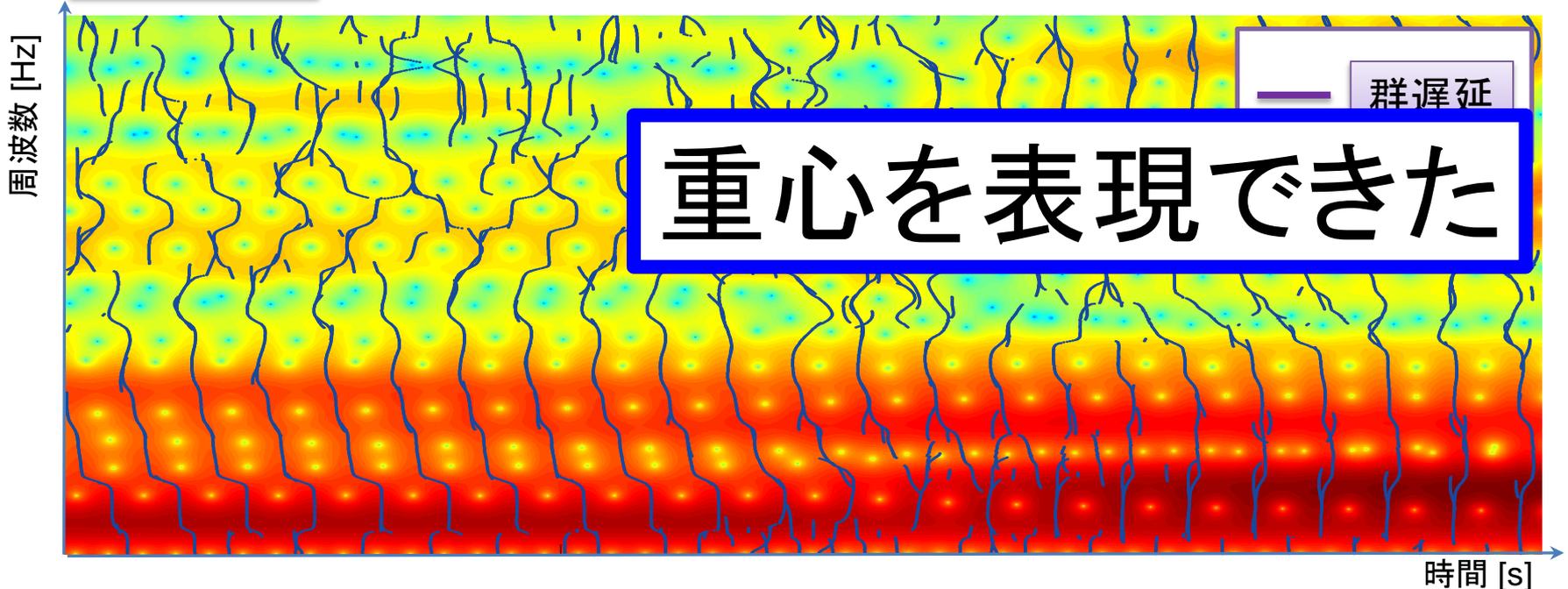
□ 各周波数での時間方向の重心



歌声波形

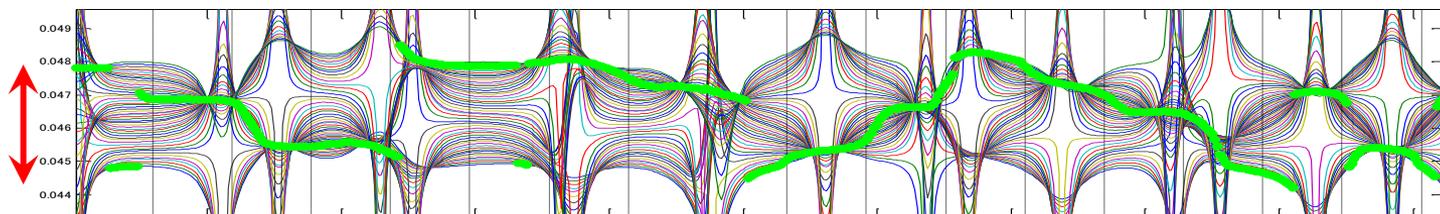


F0適応スペクトル



推定された群遅延から合成へ

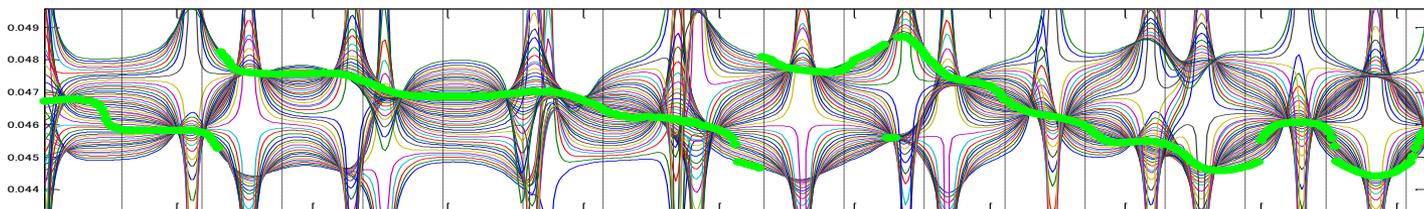
□ 基本周期に応じた時間軸の拡がりを持つ



■ 拡がりを基本周期で正規化して扱う

- 合成時には、合成の基本周期に応じて再度拡げる

□ 分析時刻に応じた相対的なずれを持つ



■ 特定の周波数帯域を0として正規化して扱う

- 現在は「 F_0 」もしくは「 $F_0 \times 1.5$ 」の値を対象

次の分析フレーム

多重フレーム統合分析:精度向上のための処理

- スペクトル包絡と同様の「問題」と「対処法」
 - 最大操作により**ステップ状の不連続性**が発生
 - 2次元ローパスフィルタにより抑制
 - **F0以下の包絡**が安定して推定できない
 - F0以下の値はF0における値で置換

- 群遅延推定における固有の「問題」と「対処法」
 - フォルマント変動などにより**群遅延が急激に変化**
 - 低域への(長い)ローパスフィルタにより抑制
 - 分析窓(ガウス窓)を合成窓(ハニング窓)に変換

[坂野 他, 2000]

多重フレーム統合分析:精度向上のための処理

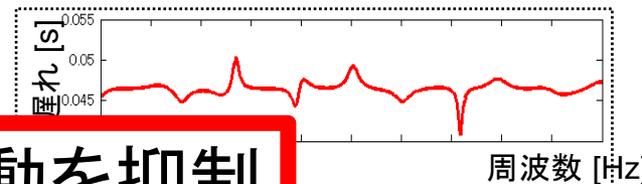
- スペクトル包絡と同様の「問題」と「対処法」
 - 最大操作により**ステップ状の不連続性**が発生
 - 2次元ローパスフィルタにより抑制
 - **F0以下の包絡**が安定して推定できない
 - F0以下の値はF0における値で置換

- 群遅延推定における固有の「問題」と「対処法」
 - フォルマント変動などにより**群遅延が急激に変化**
 - 低域への(長い)ローパスフィルタにより抑制
 - 分析窓(ガウス窓)を合成窓(ハニング窓)に変換

[坂野 他, 2000]

(再掲) 群遅延とは

各周波数での時間方向の重心

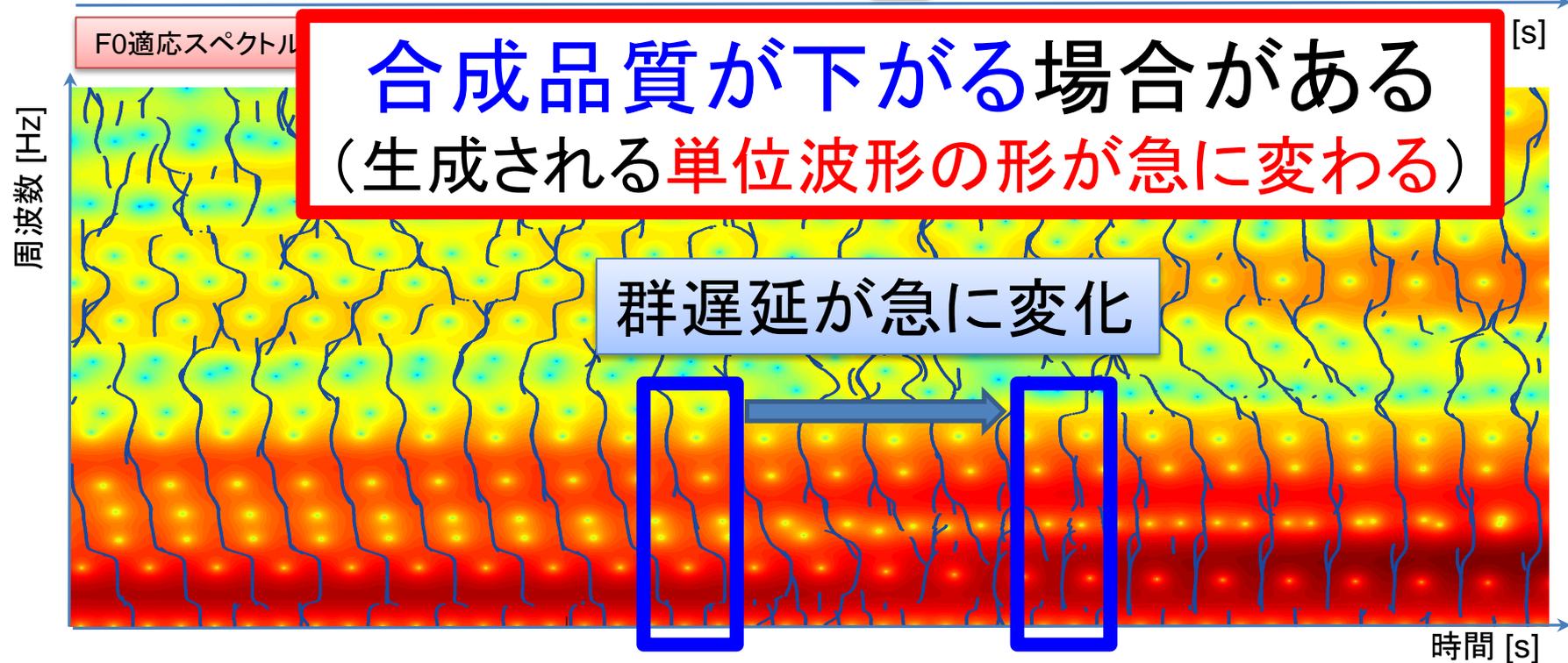


歌声波形

LPFで(低域の)変動を抑制



合成品質が下がる場合がある
(生成される単位波形の形が急に変わる)



多重フレーム統合分析:精度向上のための処理

□ スペクトル包絡と同様の処理

- 最大操作により**ステップ状の不連続性**が発生
 - 2次元**ローパスフィルタ**により抑制

50以下の包絡が内向して推定できない

【問題】

基本周期に起因する不連続性により
そのままでは扱えない

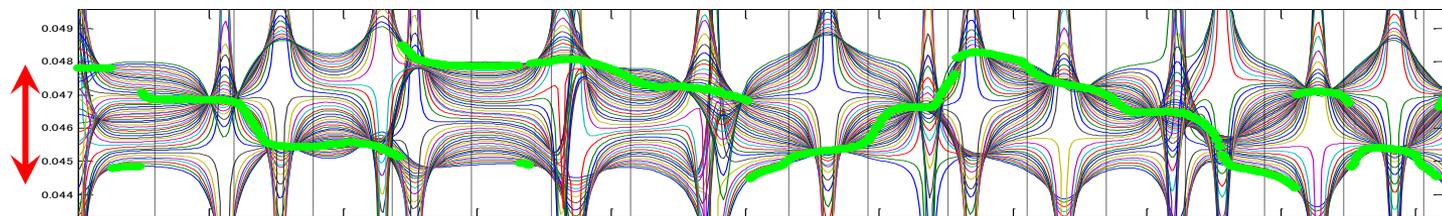
□ 群

- フォルマント変動などにより**群遅延が急激に変化**
 - 低域への(長い)**ローパスフィルタ**により抑制
- 分析窓(ガウス窓)を合成窓(ハニング窓)に変換

[坂野 他, 2000]

(本提案の)群遅延を平滑化における**問題**

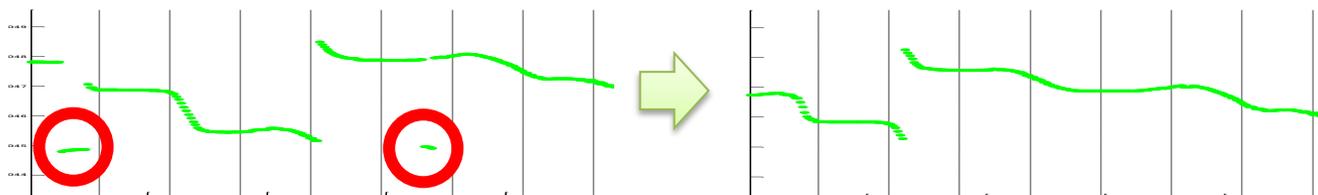
□ **基本周期**に起因する不連続性が存在する



- **sin, cos関数で展開**して、この不連続を消去する
 - **(0~基本周期)の変動**を $(-\pi \sim \pi)$ に変換して処理する
- LPF処理の後、 **\tan^{-1} 関数**で元に戻す

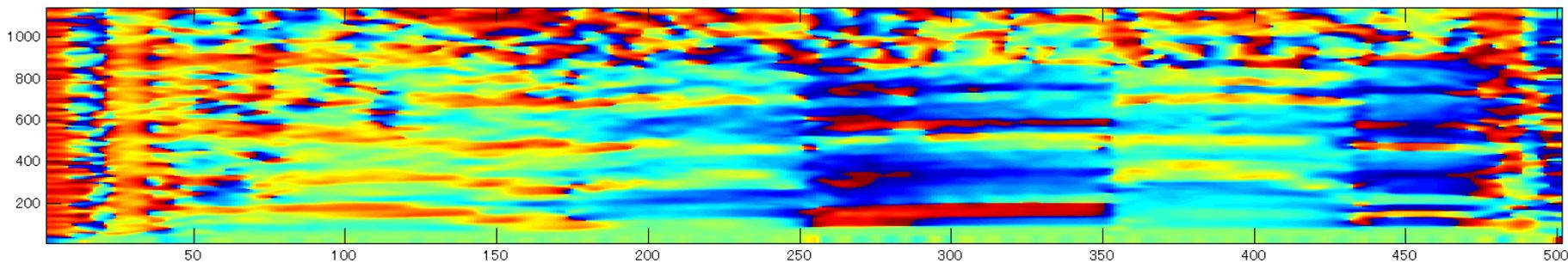
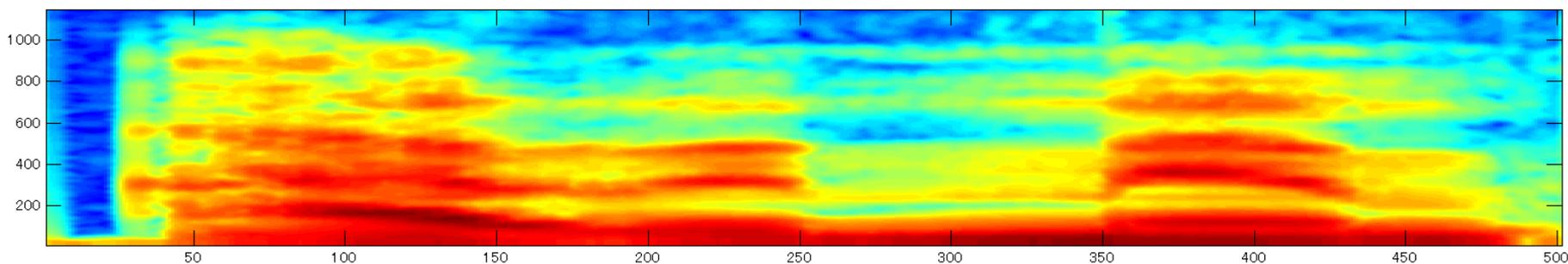
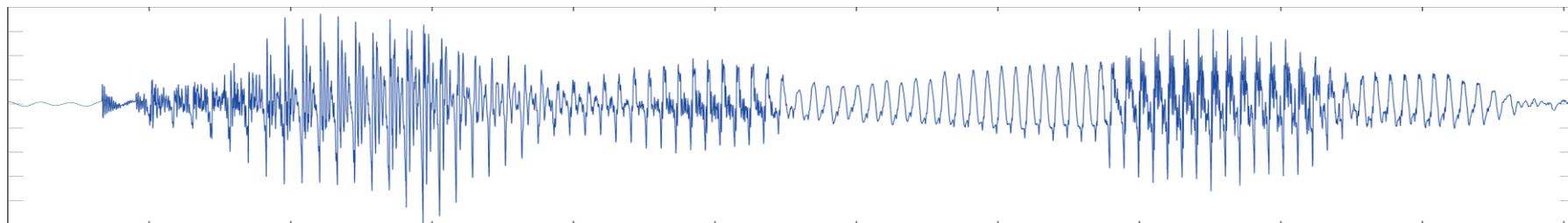
□ **合成時**

- 低域が連続的になるように**共通のオフセット**で回転



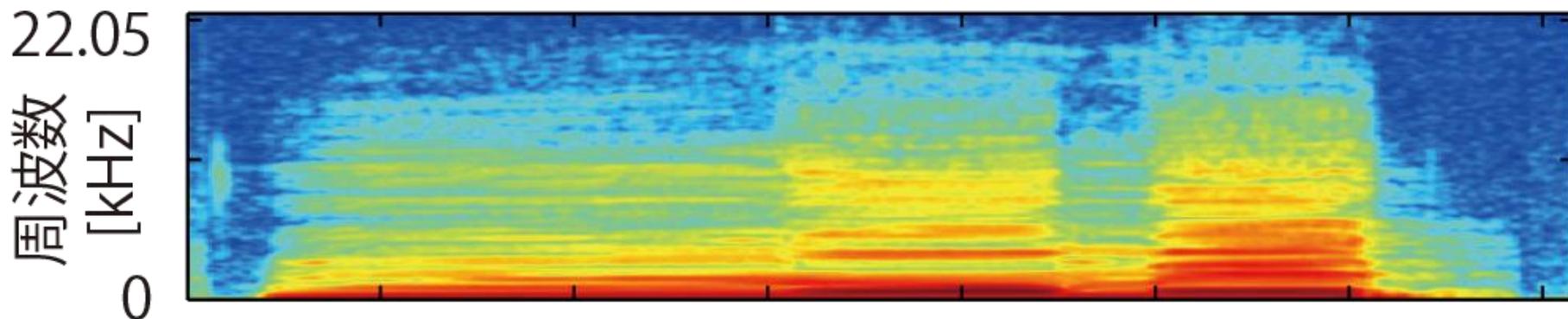
分析結果・合成結果

女性話声（分析結果は低域を表示）

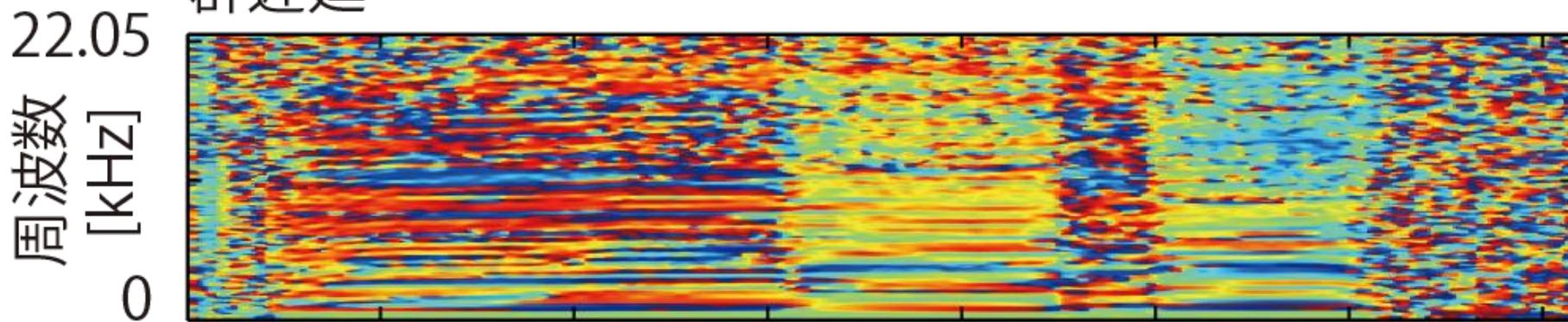


歌声の例(分析結果)

スペクトル包絡



群遅延

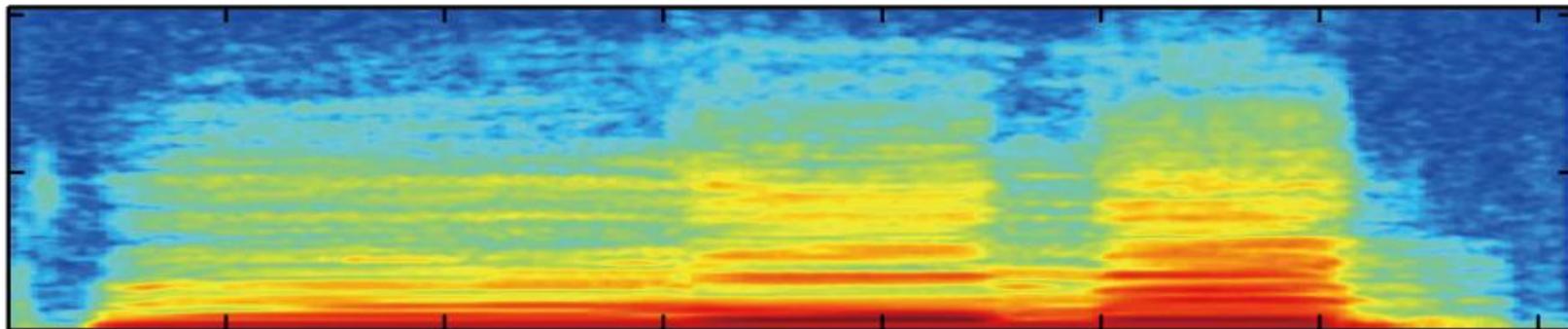


歌声の例（再合成結果）：音高・時間軸も変更可能

スペクトル包絡

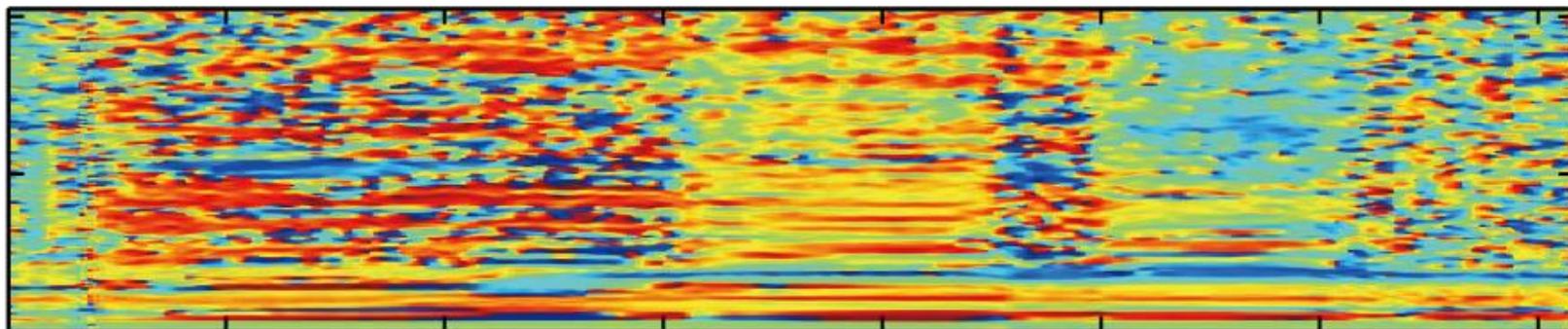


22.05
周波数 [kHz]
0



群遅延

22.05
周波数 [kHz]
0



まとめ

- **F0適応多重フレーム統合分析手法を提案**
 - **対数スペクトル距離を既存手法と比較**
 - 13サンプル(14サンプル中)でいずれかより低い
 - 8サンプル(14サンプル中)でどちらよりも低い
 - **ピッチマークや音素ラベルなしで推定可能**
 - **ピッチマーク分析の限界を超えることができる**

- **今後の課題：手法をより洗練させる**
 - LPF処理を排除、群遅延の扱いを改善
 - 音素モデルの利用や確率モデルへの発展、等

まとめ

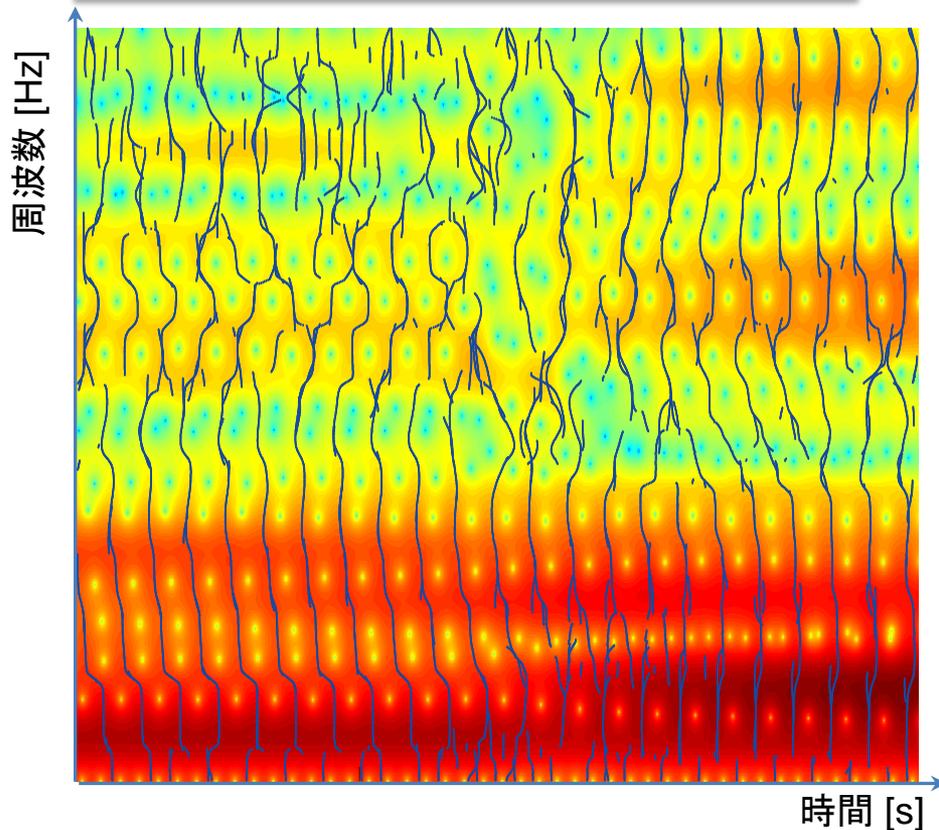
- F0適応多重フレーム統合分析手法を提案
 - 対数スペクトル距離を既存手法と比較
 - 13サンプル(14サンプル中)でいずれかより低い
 - 8サンプル(14サンプル中)でどちらよりも低い
 - ピッチマークや音素ラベルなしで推定可能
 - ピッチマーク分析の限界を超えることができる

- 今後の課題：手法をより洗練させる
 - LPF処理を排除、群遅延の扱いを改善
 - 音素モデルの利用や確率モデルへの発展、等

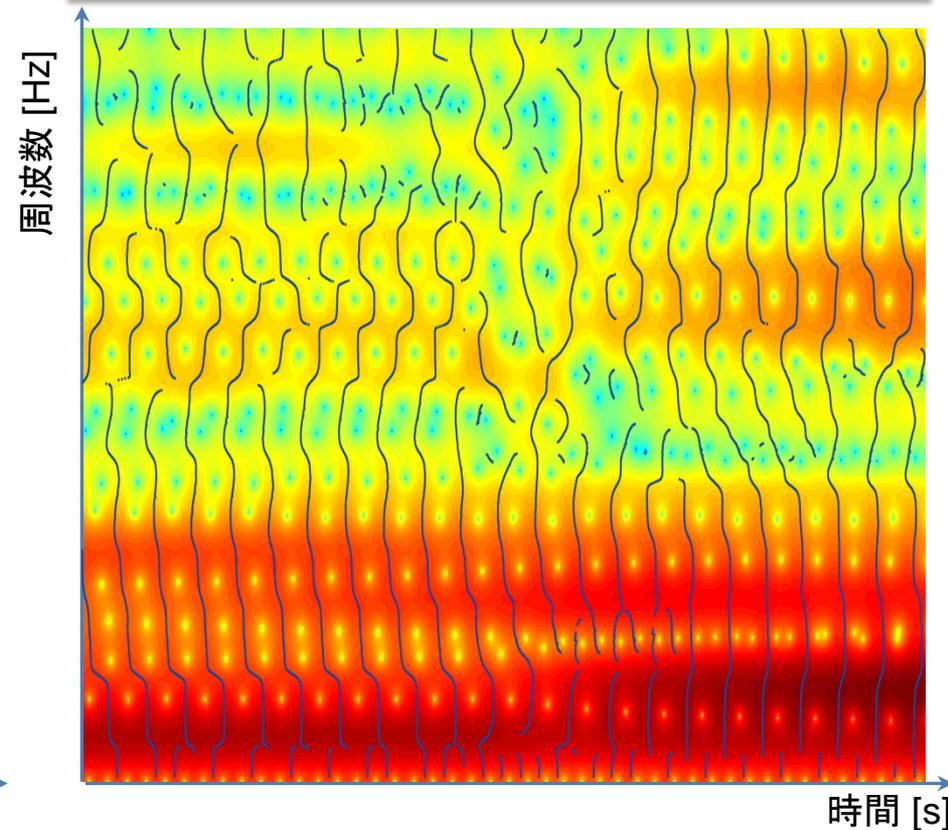
手法の洗練

最大包絡線の選択でピーク検出を行う

F0適応スペクトルと群遅延(本提案)



F0適応スペクトルと群遅延(ピーク検出)

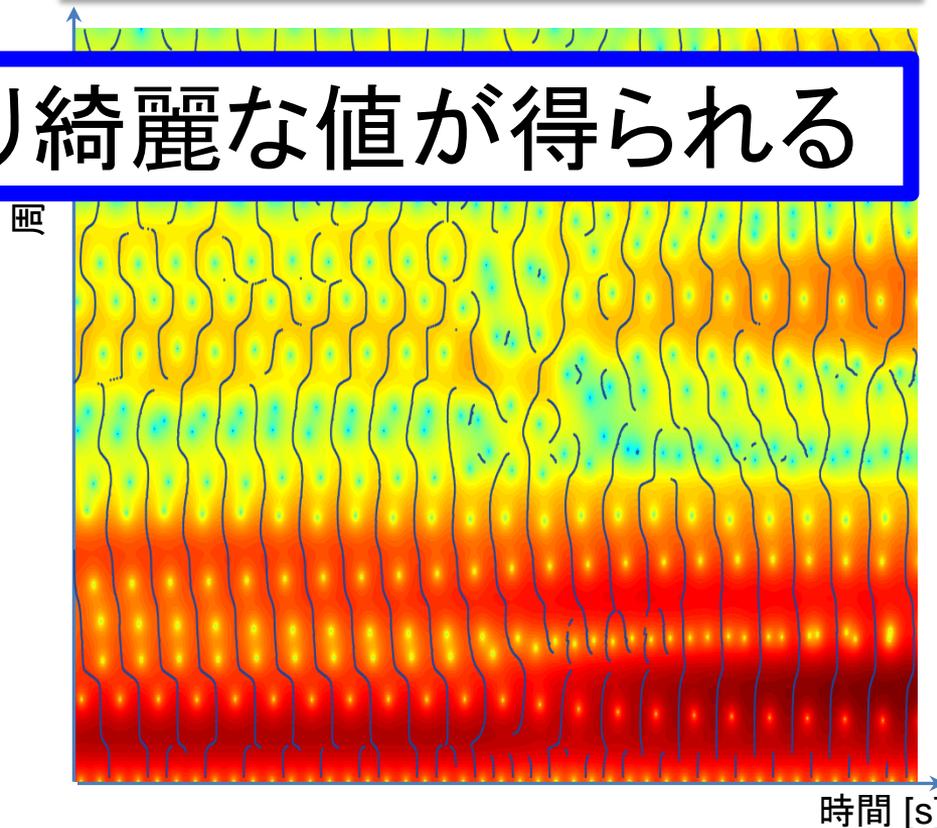
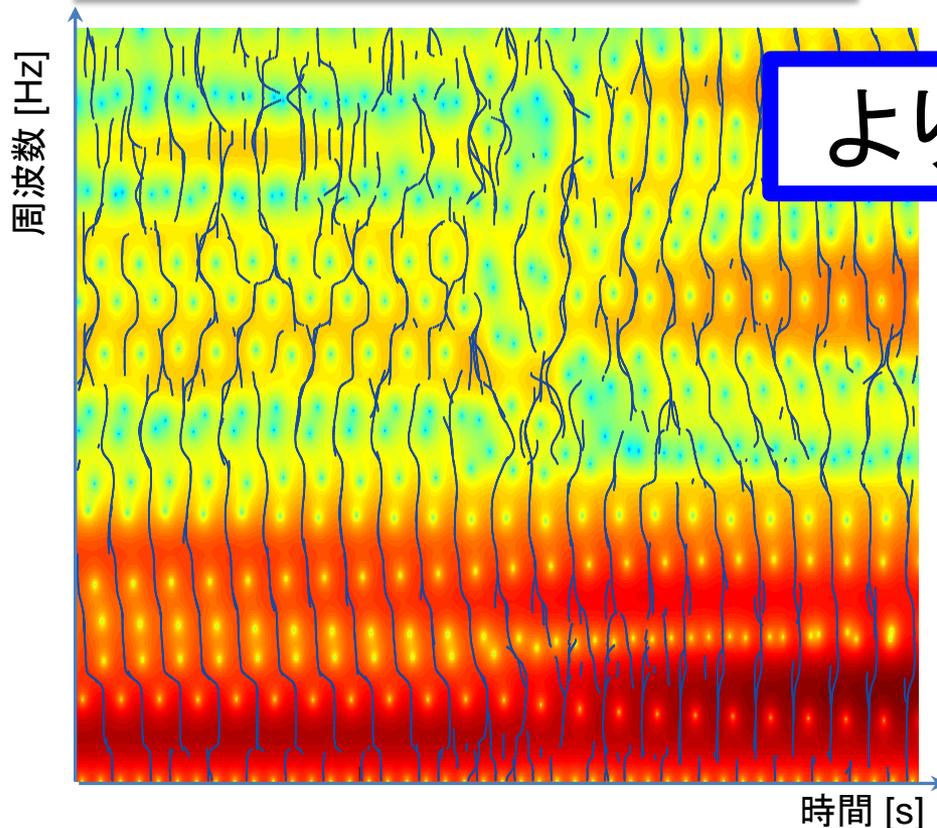


手法の洗練

最大包絡線の選択でピーク検出を行う

F0適応スペクトルと群遅延(本提案)

F0適応スペクトルと群遅延(ピーク検出)



より綺麗な値が得られる

まとめ

- **F0適応多重フレーム統合分析手法を提案**
 - **対数スペクトル距離を既存手法と比較**
 - 13サンプル(14サンプル中)でいずれかより低い
 - 8サンプル(14サンプル中)でどちらよりも低い
 - **ピッチマークや音素ラベルなしで推定可能**
 - **ピッチマーク分析の限界を超えることができる**

- **今後の課題：手法をより洗練させる**
 - LPF処理を排除、群遅延の扱いを改善
 - 音素モデルの利用や確率モデルへの発展、等