VocaListener

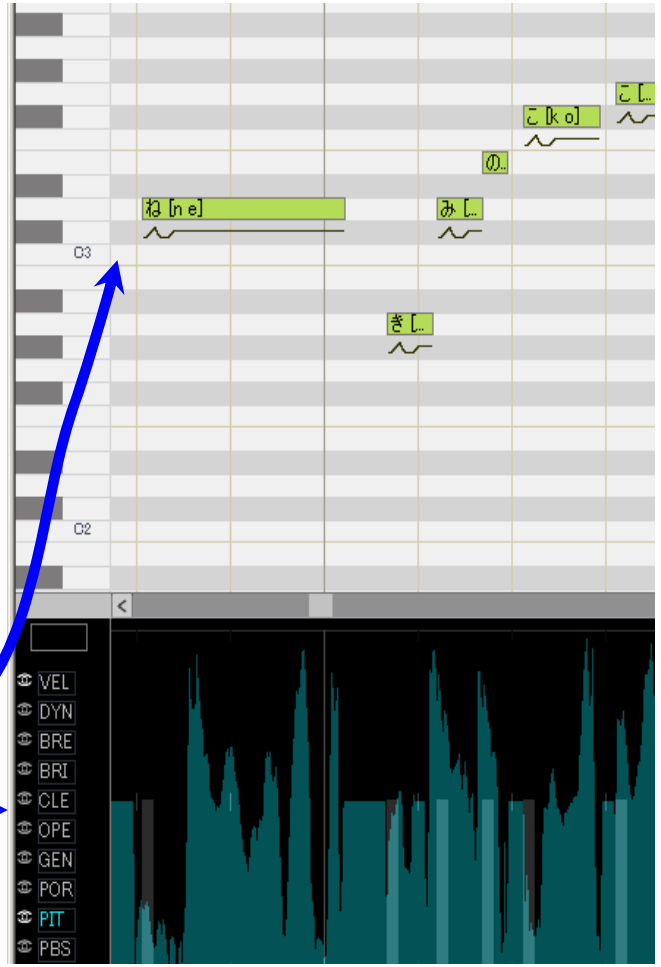
ユーザ歌唱を真似る歌声合成パラメータを自動推定するシステムの提案

中野倫靖, 後藤真孝
（産業技術総合研究所）

2008年5月28日
第75回音楽情報科学研究会(SIGMUS)
第128回ヒューマンコンピュータインタラクション研究会 (SIGHCI)

現状の歌声合成の使い方

- 歌声合成システムを選択
  [ ] Vocaloid [✓] Vocaloid2
- 音源を選択
  [✓] 初音ミク（CV01）
  [ ] 鏡音リン（CV02）
- 歌声合成パラメータの入力（打ち込み）
  - 楽譜情報，歌詞情報
  - 表情パラメータ（歌唱スタイル）
- 合成歌唱を得る



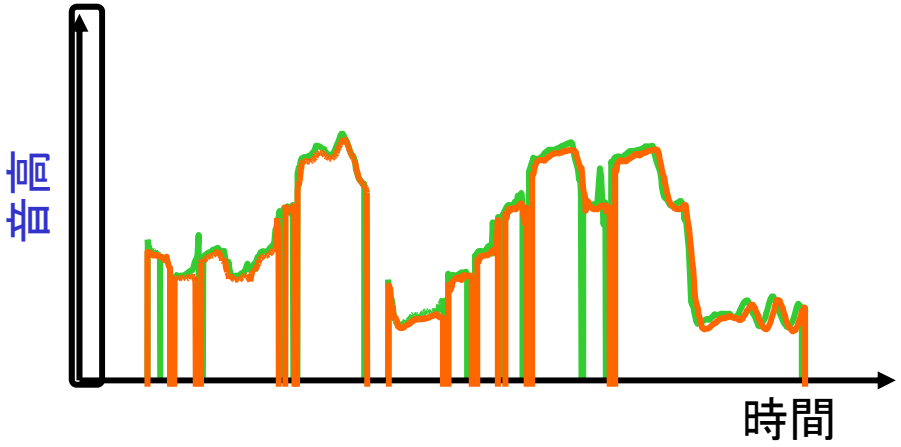
例: Vocaloid Editor

問題点

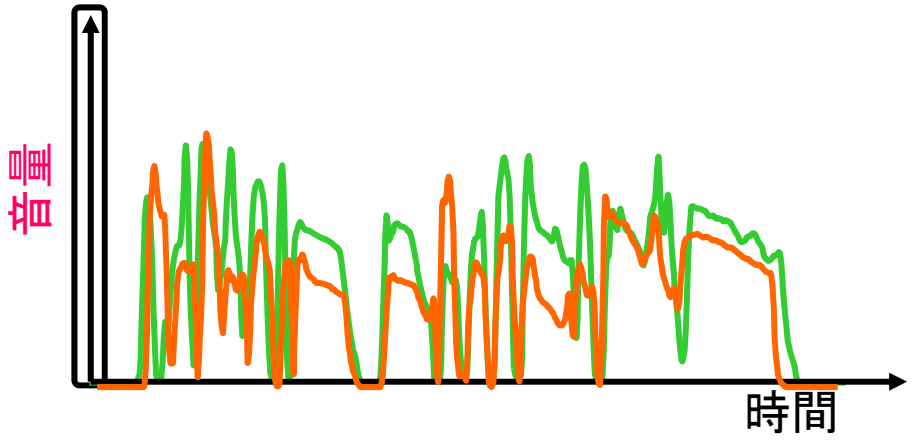
- 音符を並べるだけでは自然性が低い場合もある
  - 歌詞:「立ち止まる時 また ふと振り返る」

CV01 CV02

音高(声の高さ)



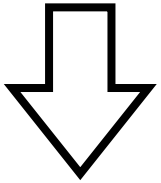
音量(声の大きさ)



- 品質を高くしようとすると
  パラメータ調整に時間がかかる

研究目的

- 高品質な歌声を手軽に合成すること
  - それによって歌声合成のユーザの支援を目指す



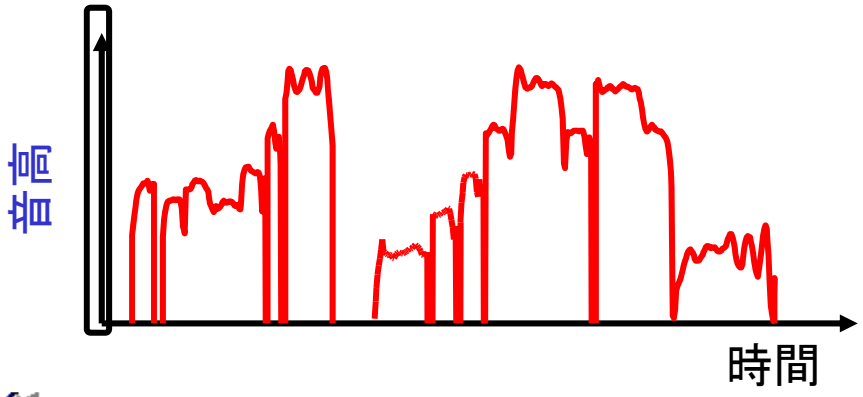
ユーザ歌唱を真似る歌声合成パラメータを自動推定する VocaListener を提案

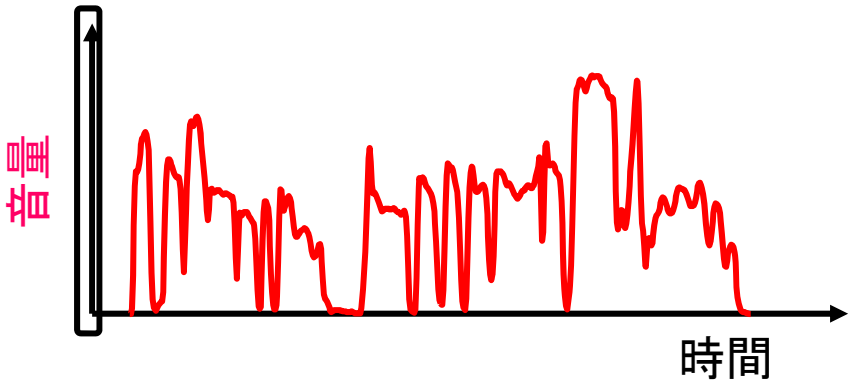
VocaListener でユーザ歌唱を真似る

音高（声の高さ）

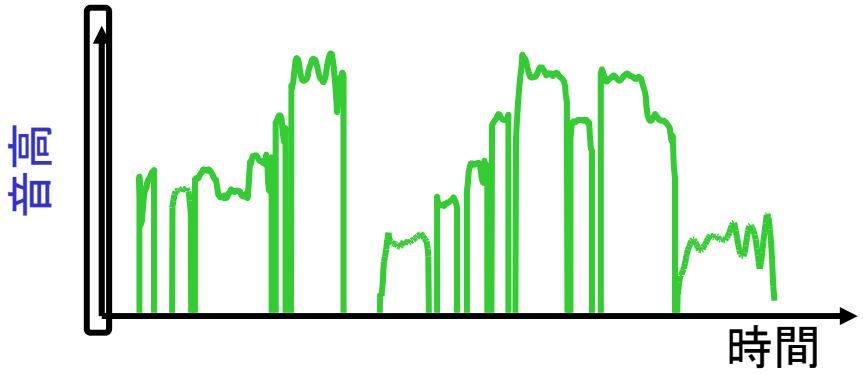
音量（声の大きさ）

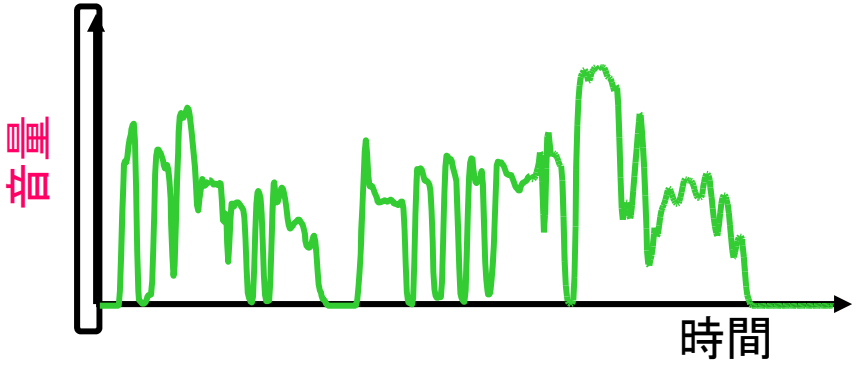
ユーザ歌唱





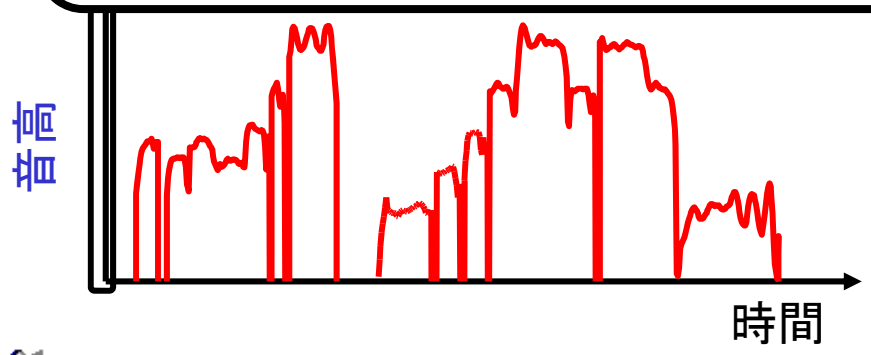
VocaListenerでパラメータ推定して合成（CV01）

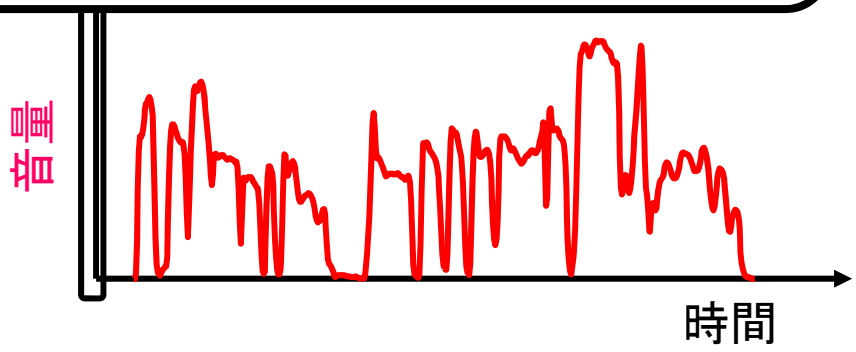




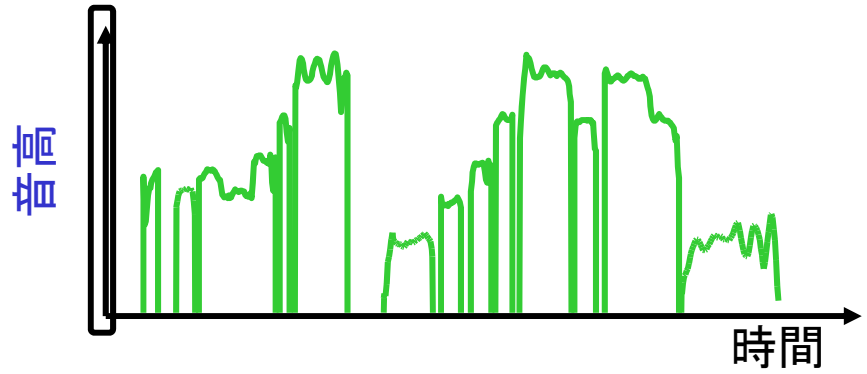
VocaListener でユーザ歌唱を真似る

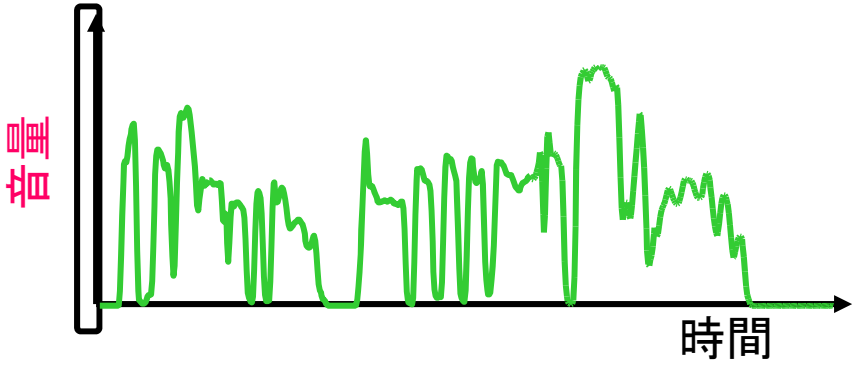
ユーザ歌唱とほぼ同じ表情（音高、音量の変化）の合成歌唱音声が得られる





VocaListenerでパラメータ推定して合成（CV01）

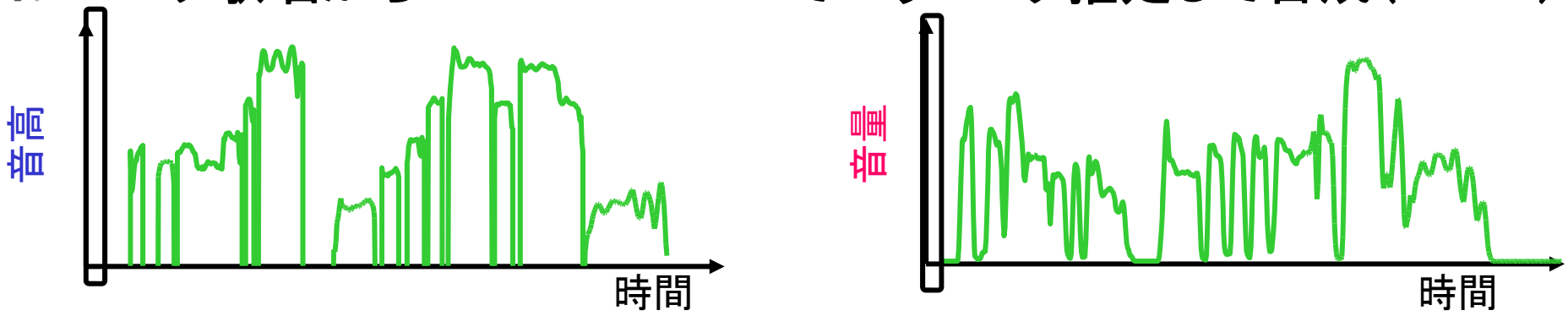




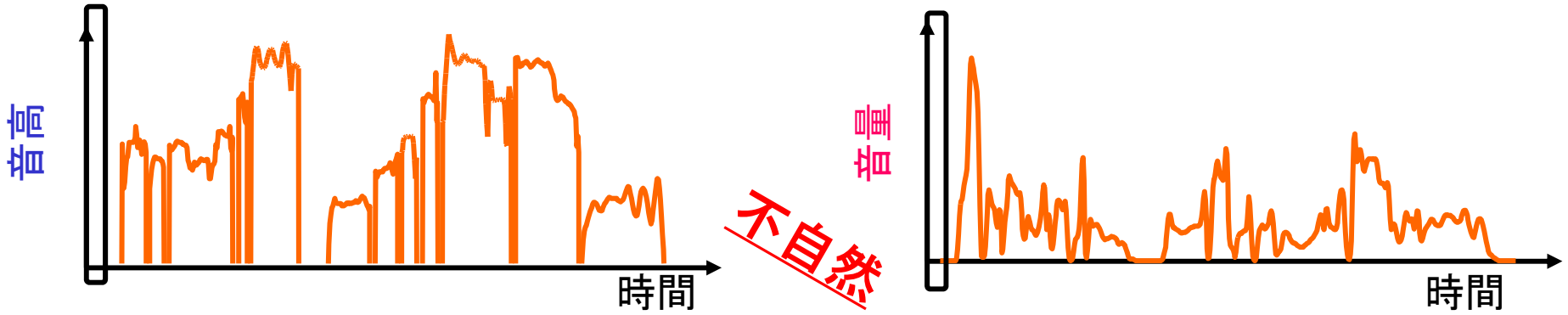
もう一つの問題点

- 同じ表情パラメータでも音源によって合成結果が変わる

ユーザ歌唱から VocaListener でパラメータ推定して合成（CV01）



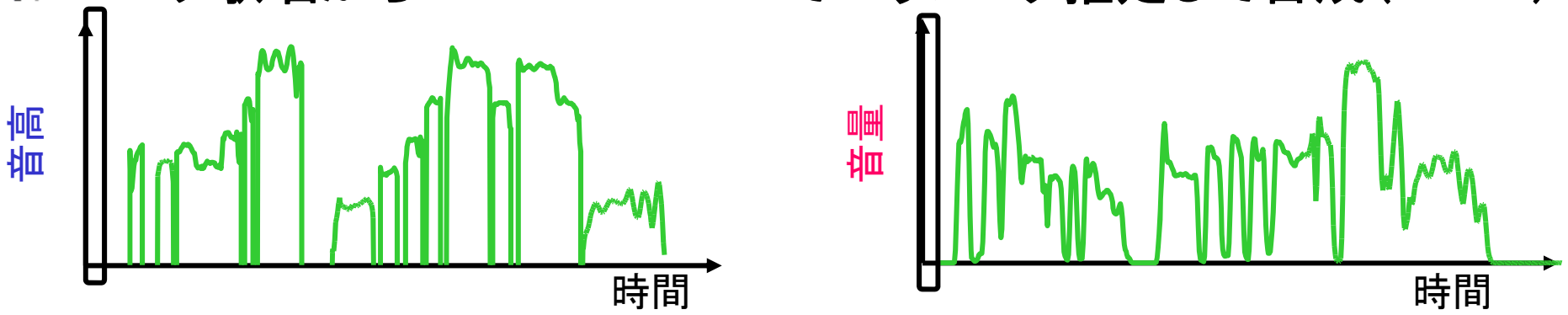
音源データを変えて同じ表情パラメータで合成（CV01 ⇒ CV02）



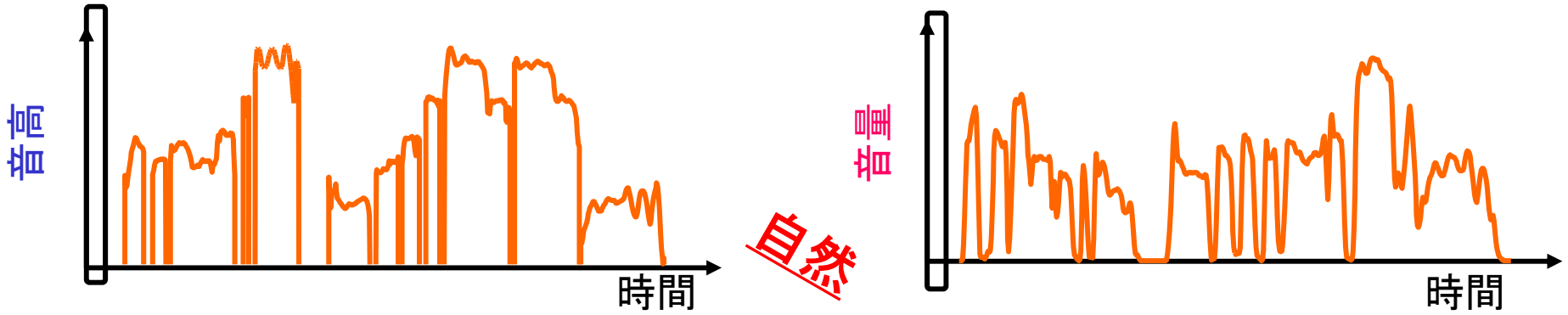
VocaListener による表情パラメータ再推定

- 同じ表情パラメータでも音源によって合成結果が変わる

ユーザ歌唱から VocaListener でパラメータ推定して合成（CV01）

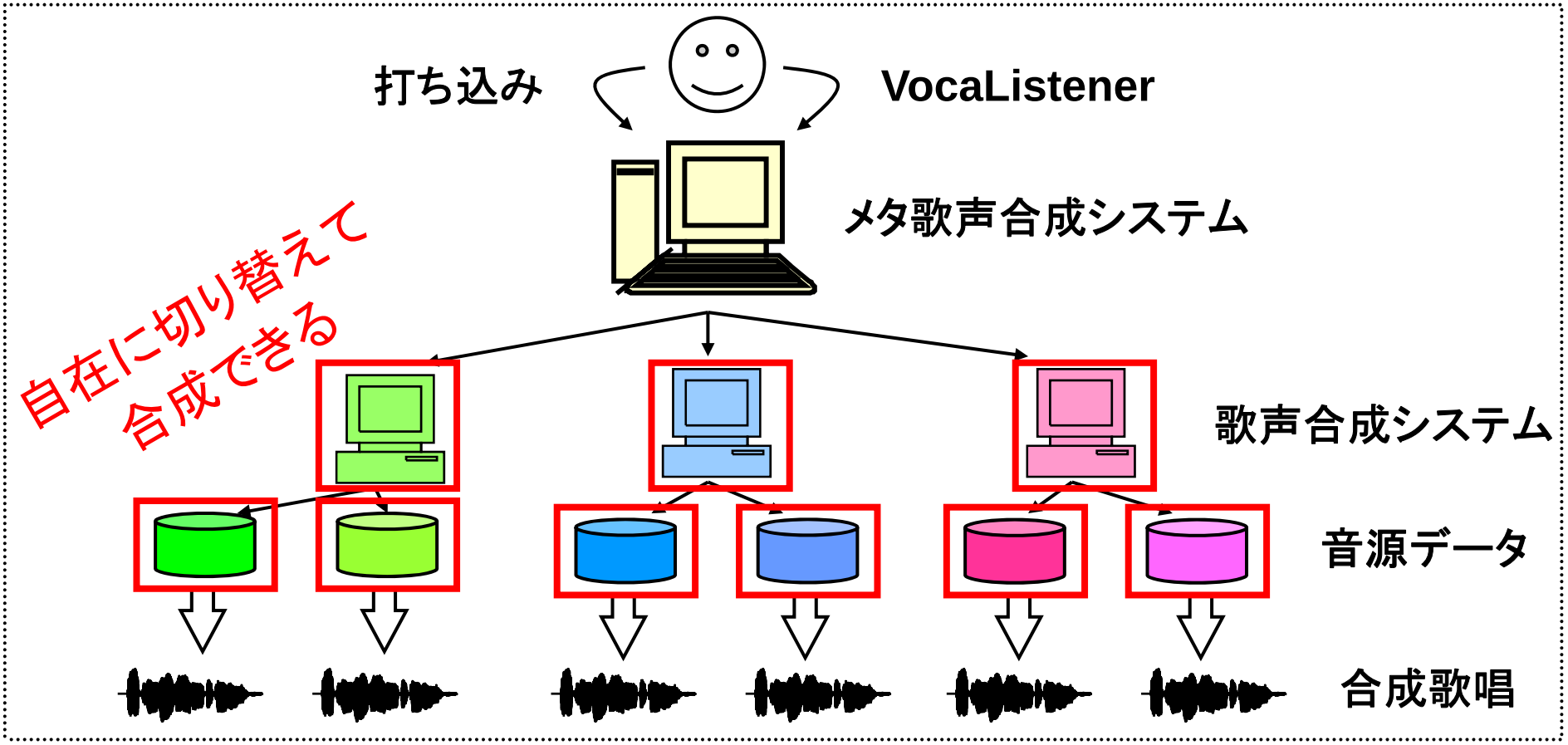


VocaListener でパラメータを再推定して合成（CV01 ⇒ CV02）



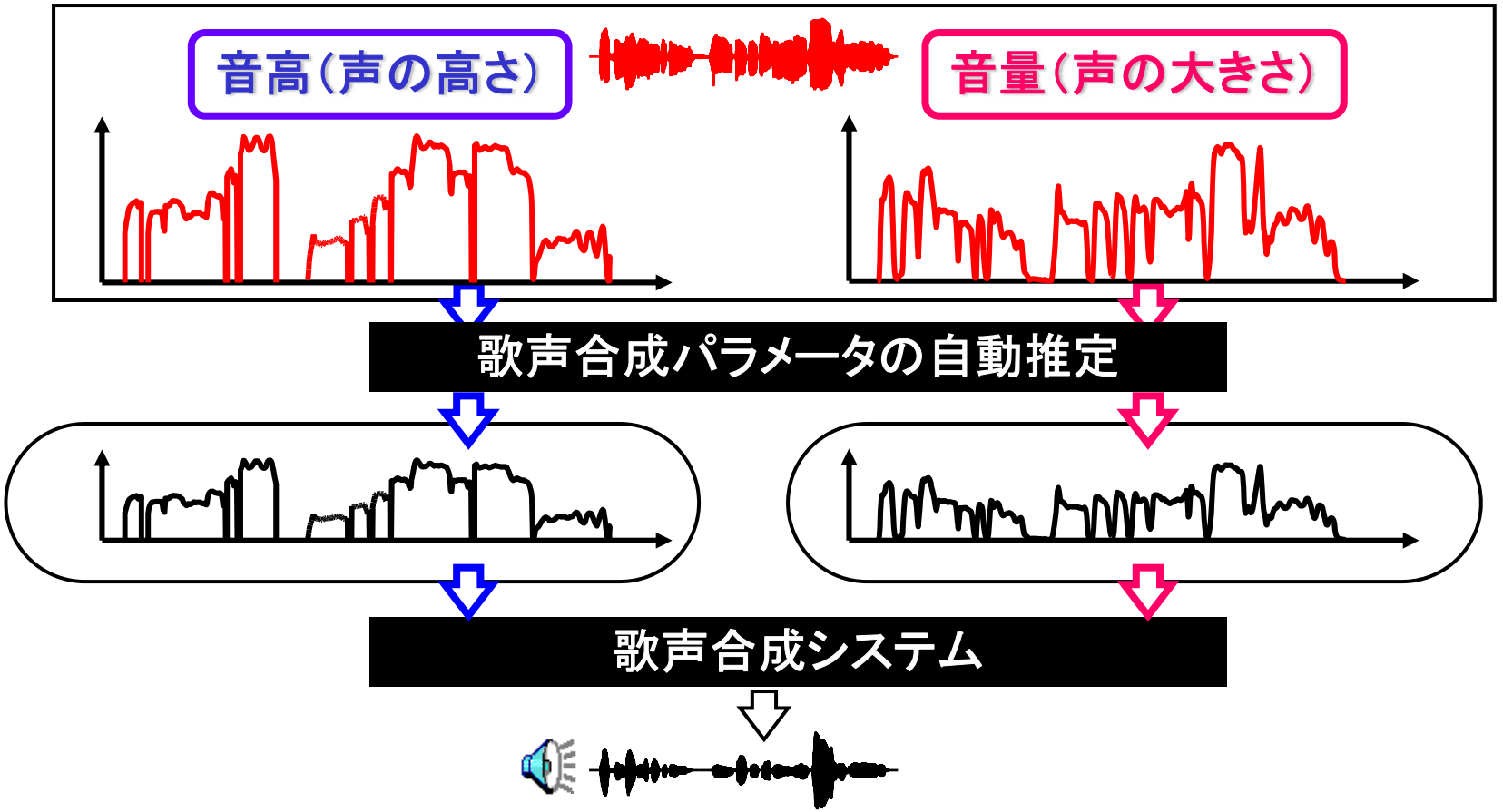
『メタ歌声合成システム』の提案

- 一度合成パラメータを調整するだけで
  様々な歌声を合成できるシステム



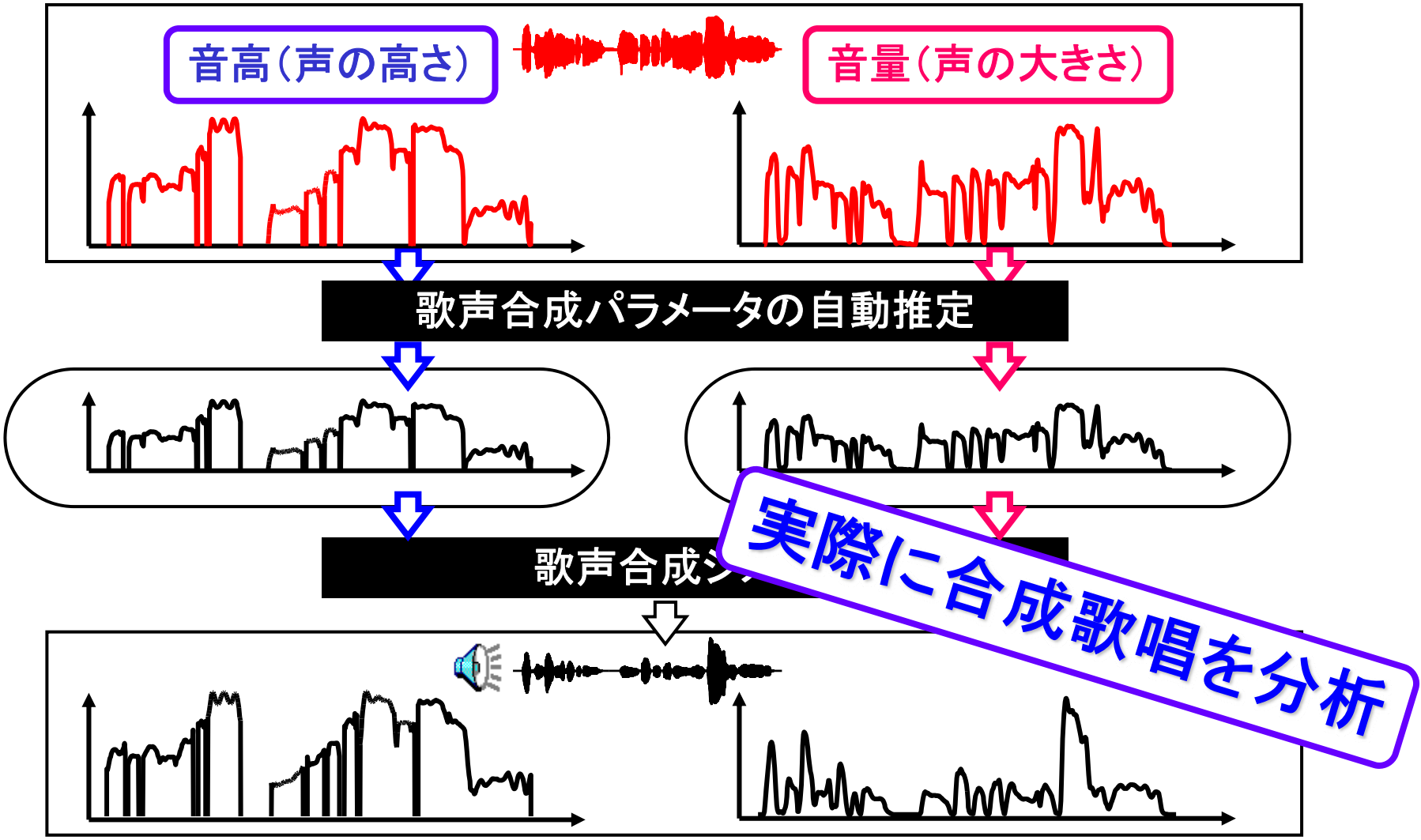
VocaListener の 実現方法

従来の歌声合成パラメータ自動推定[Janer *et al.*, 2006]



歌唱音声を分析して音高・音量を推定し
それを合成パラメータへ変換して与えていた

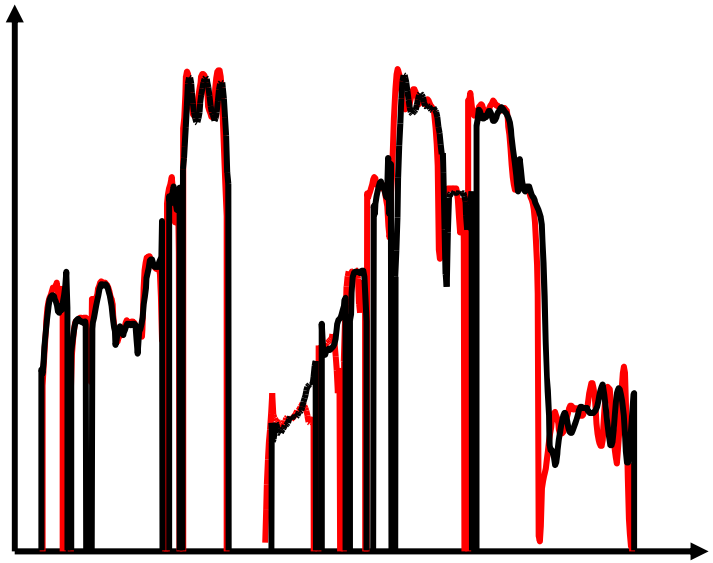
従来の問題点(1)



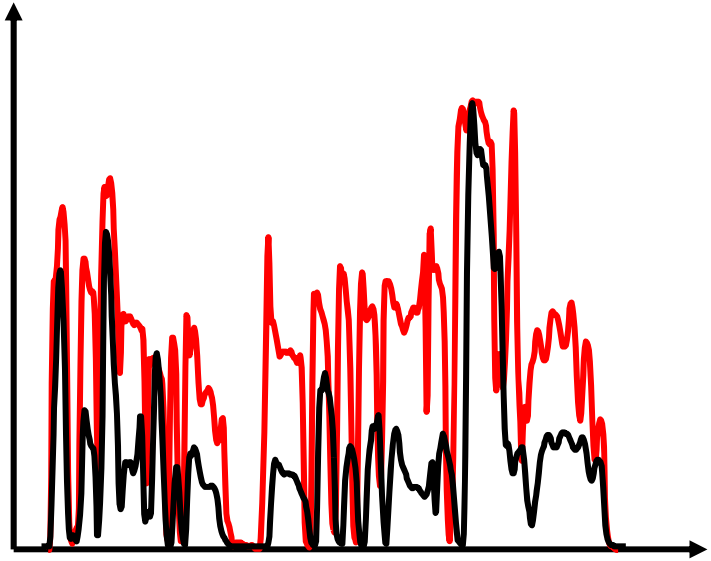
従来の問題点(1)

与えたパラメータ通りに合成されていない

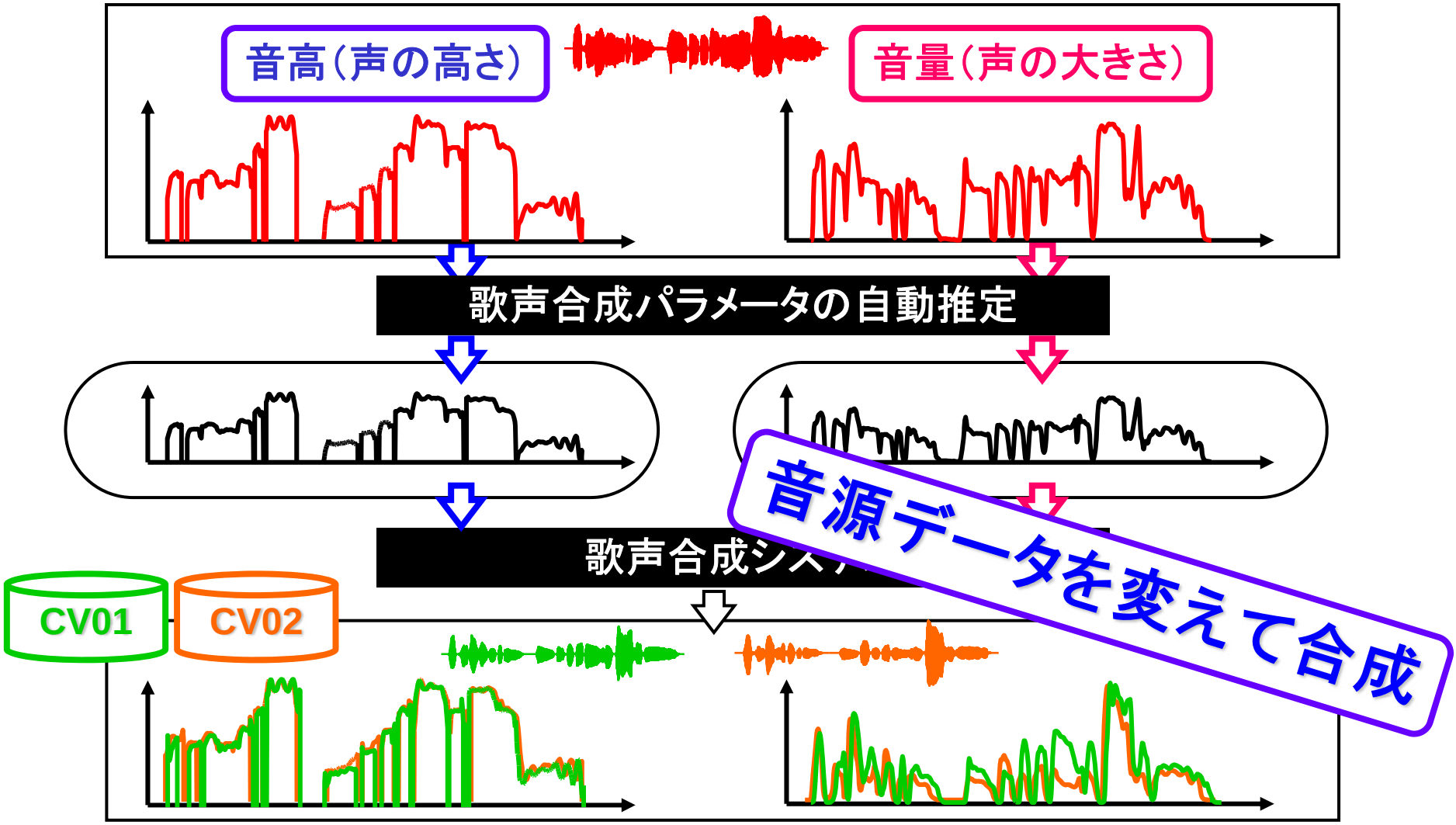
音高（声の高さ）



音量（声の大きさ）



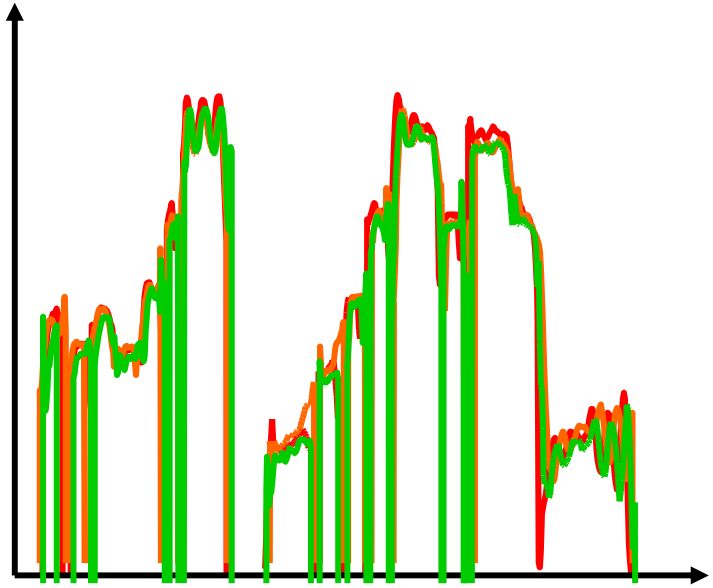
従来の問題点(2)


音高（声の高さ）
音量（声の大きさ）
歌声合成パラメータの自動推定
歌声合成シス
CV01
CV02
音源データを変えて合成

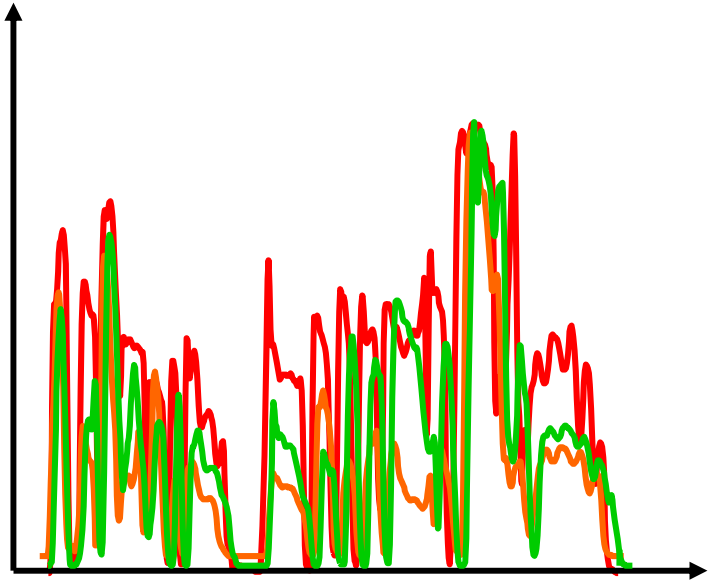
従来の問題点(2)

音源データが異なると
さらに違う合成結果になる

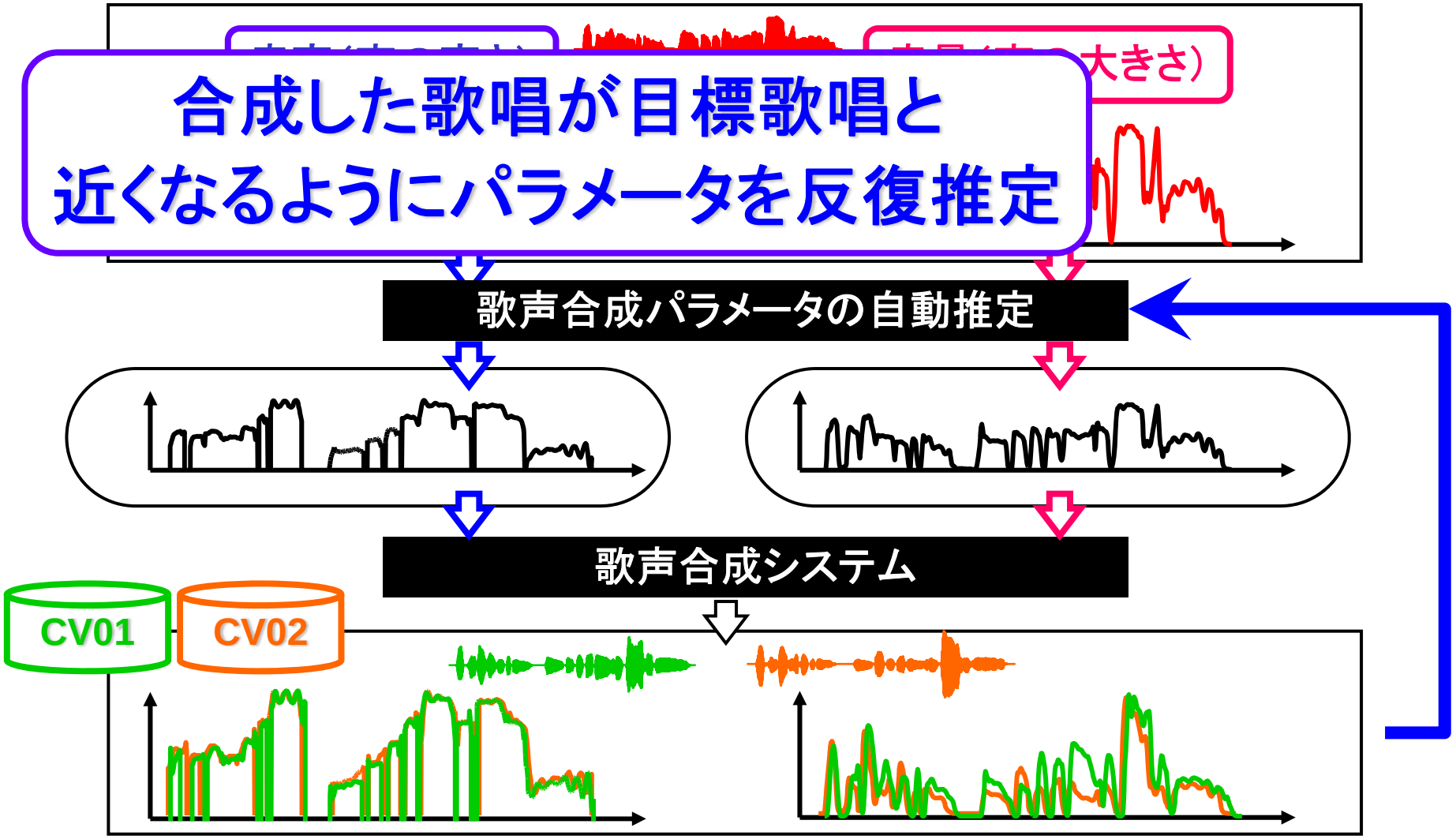
音高（声の高さ）



音量（声の大きさ）

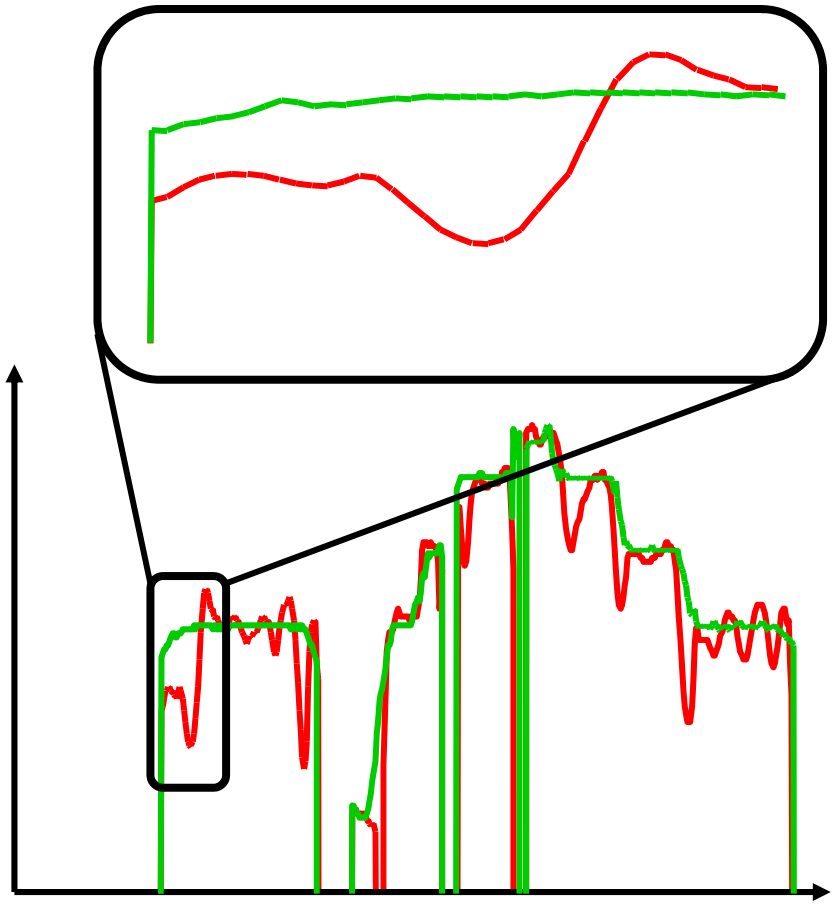


本研究の解決法

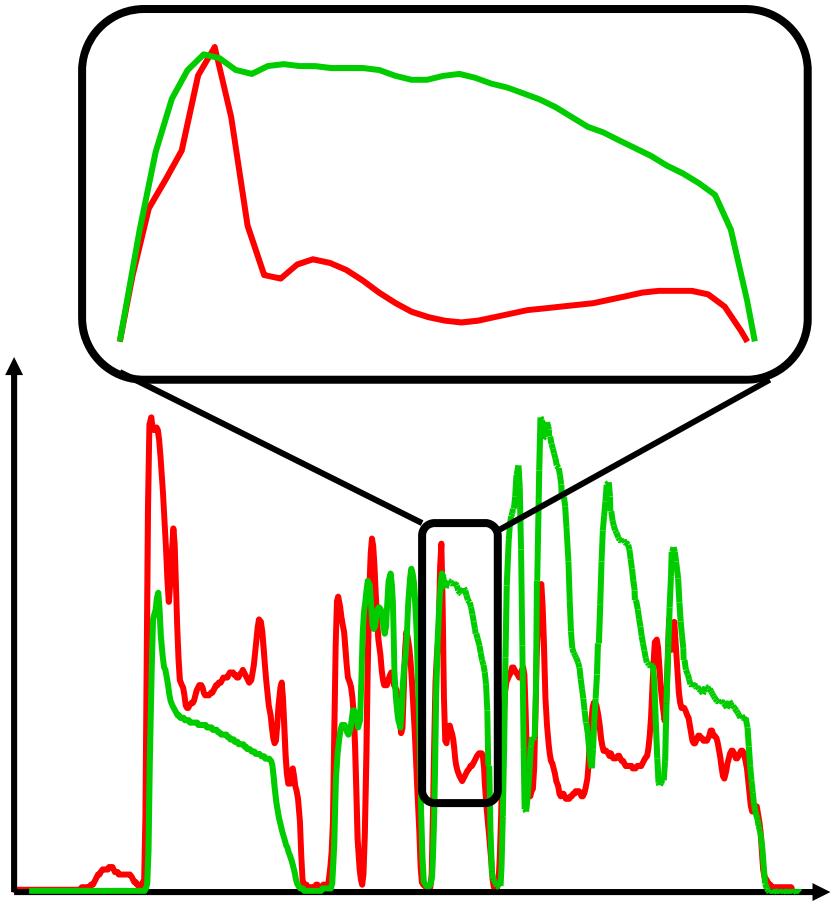

大きさ）
合成した歌唱が目標歌唱と
近くなるようにパラメータを反復推定
歌声合成パラメータの自動推定
歌声合成システム
CV01
CV02

反復推定による音高・音量の収束（初期値）

音高（声の高さ）



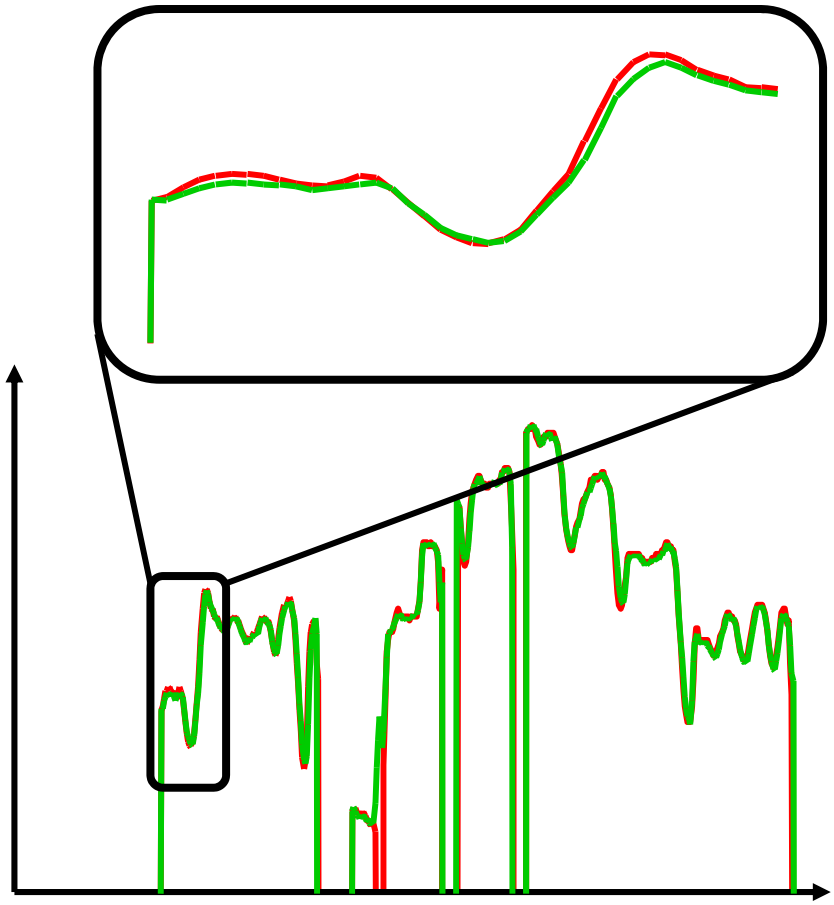
音量（声の大きさ）

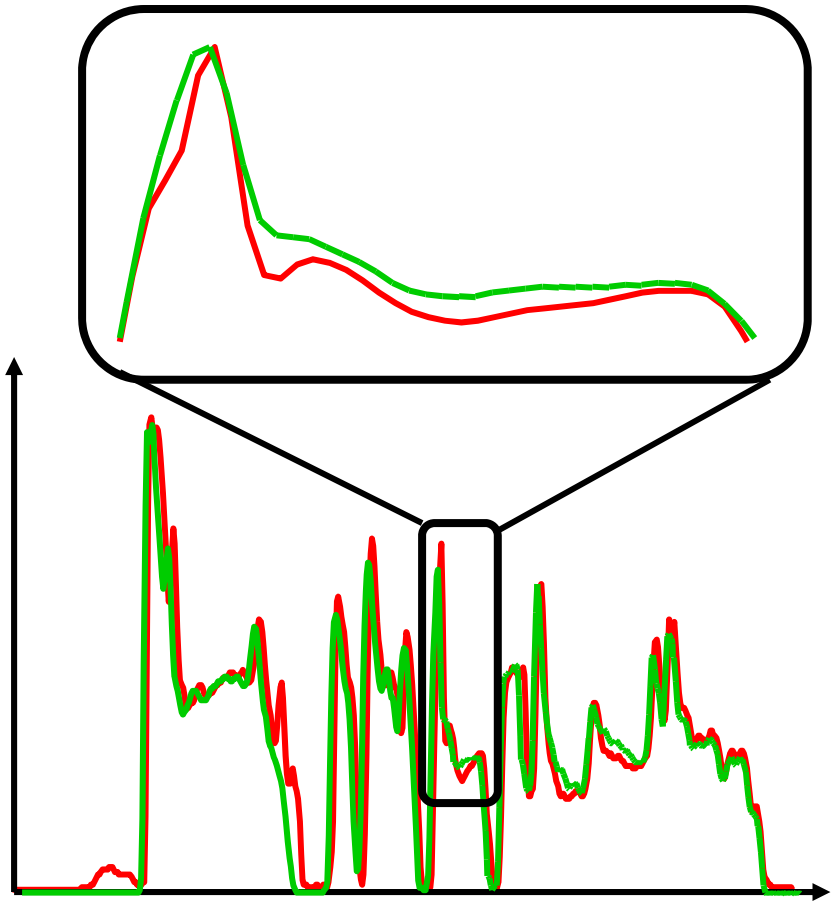


反復推定による音高・音量の収束（反復1回）

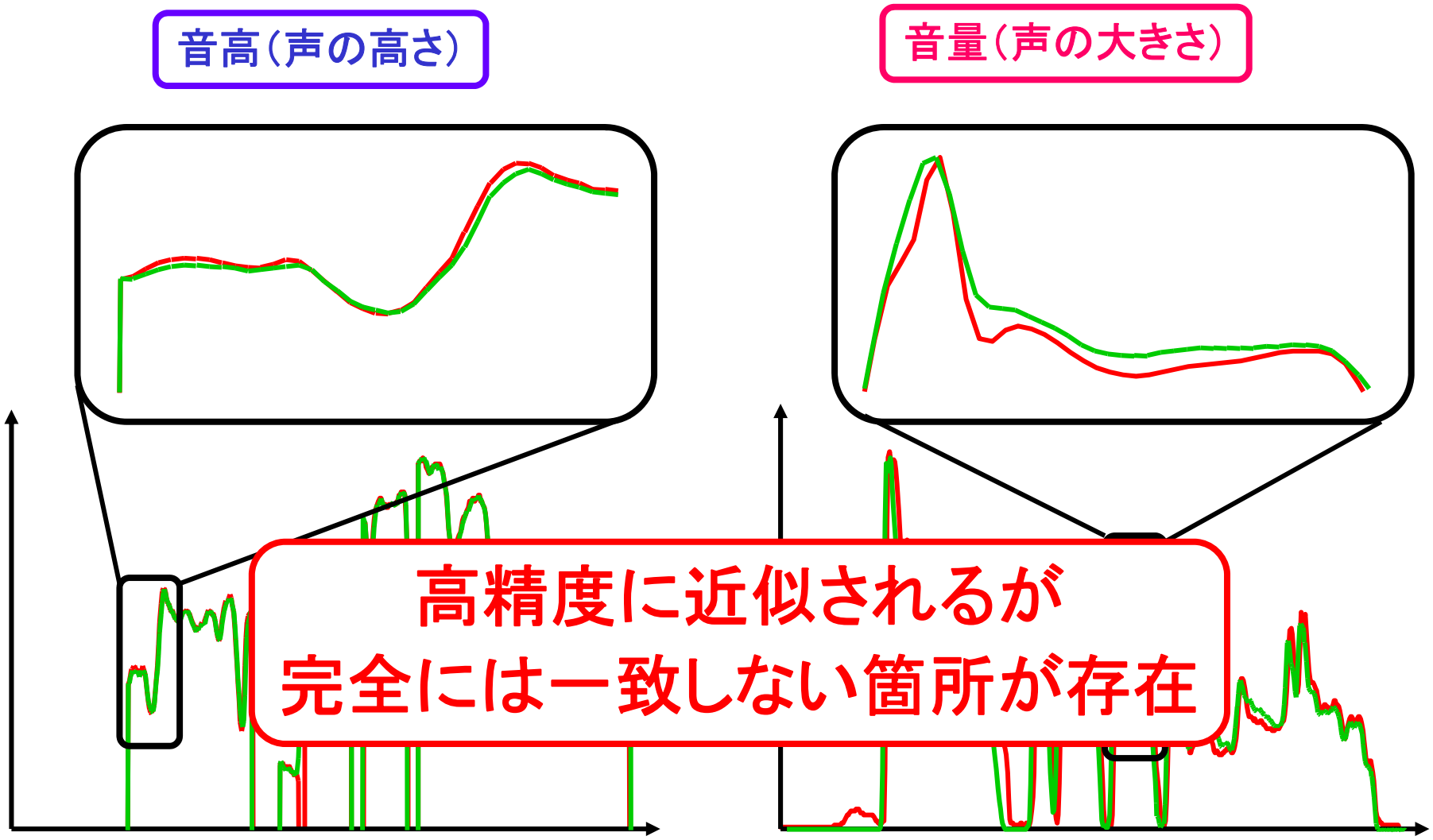
音高（声の高さ）

音量（声の大きさ）





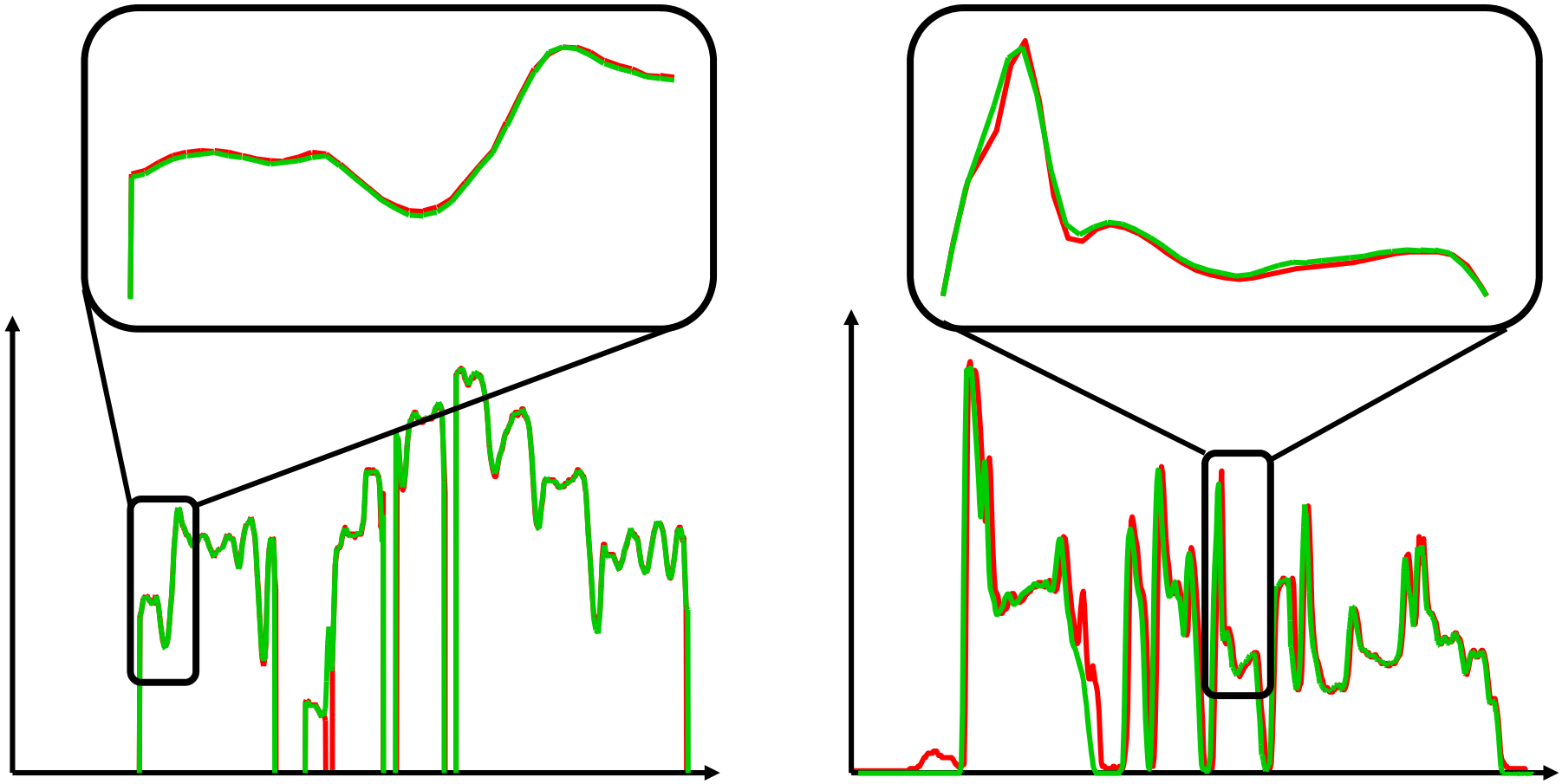
反復推定による音高・音量の収束（反復1回）



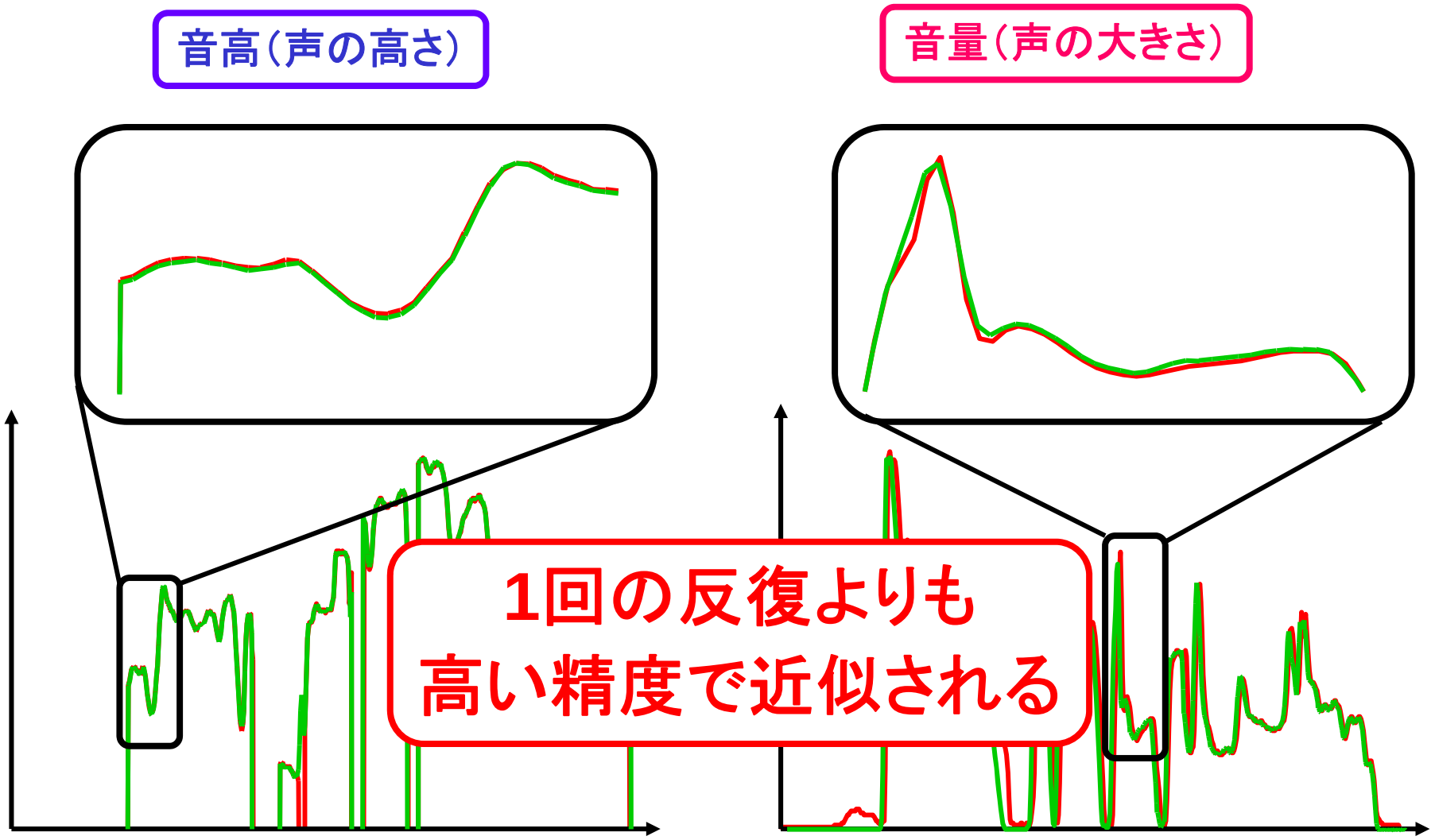
反復推定による音高・音量の収束（反復2回）

音高（声の高さ）

音量（声の大きさ）

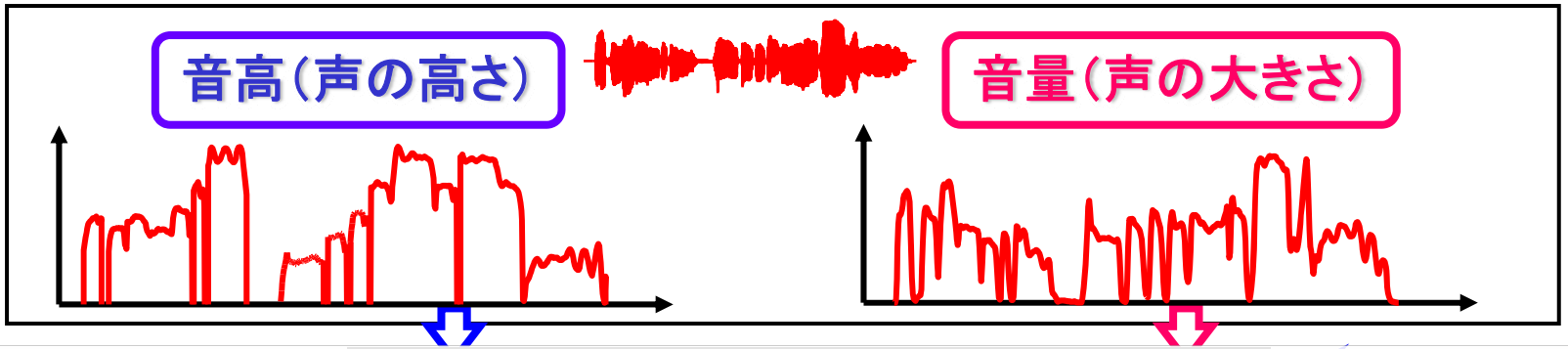


反復推定による音高・音量の収束（反復2回）



VocaListener 処理概要

目標歌唱の分析



歌声合成パラメータの自動推定

既存の手法を利用して目標歌唱を分析

音高：SWIPE で基本周波数を推定

音量：音声波形の振幅の実効値を計算

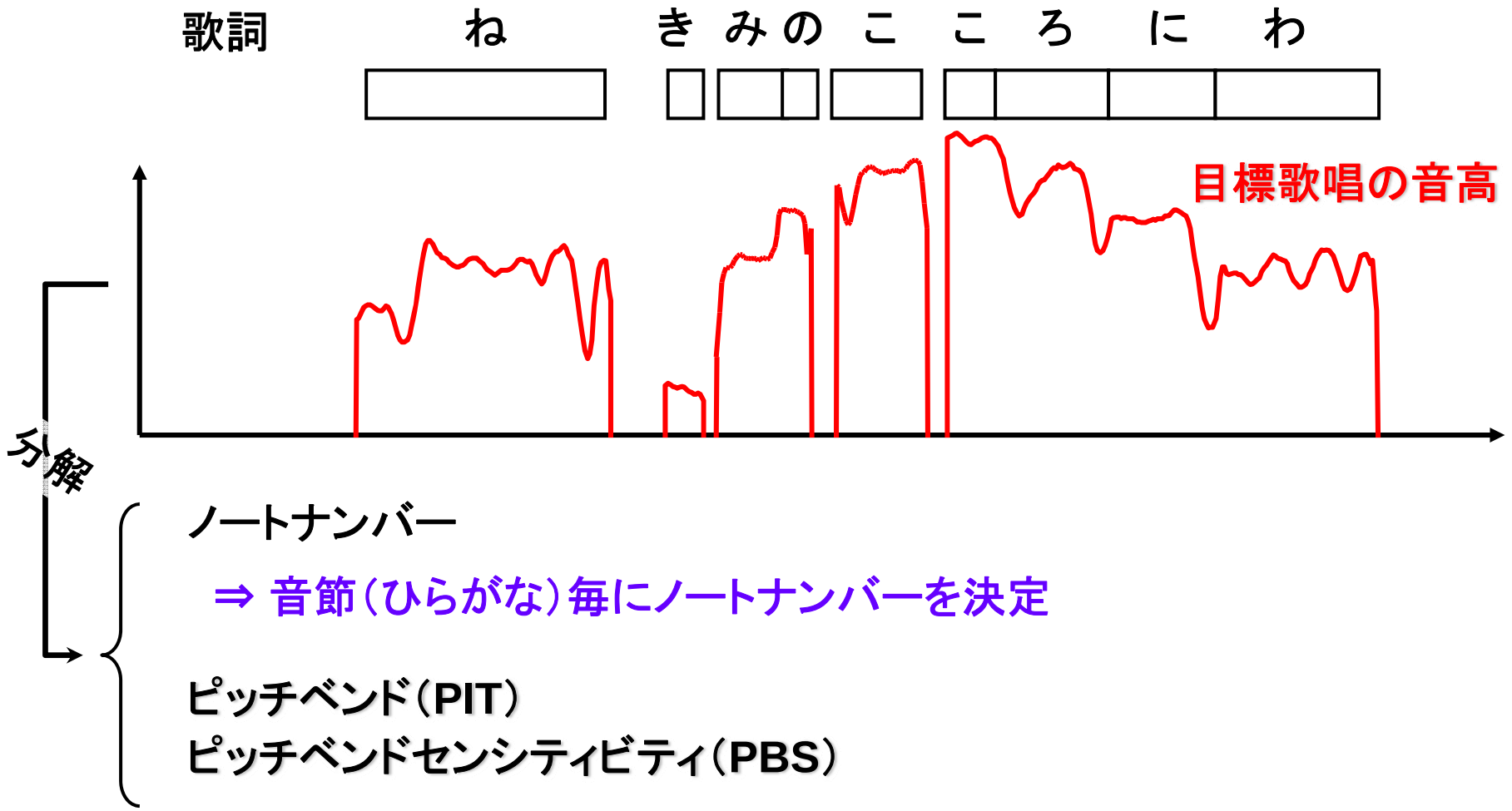
歌声合成パラメータ(音高パラメータ)

音高パラメータ:Vocaloid2 の場合

- ノートナンバー → 各音節(音符)の音高に相当
- ピッチベンド(PIT)
- ピッチベンドセンシティビティ(PBS)

→ ノートナンバーからの相対音高を表現

音高から音高パラメータへ



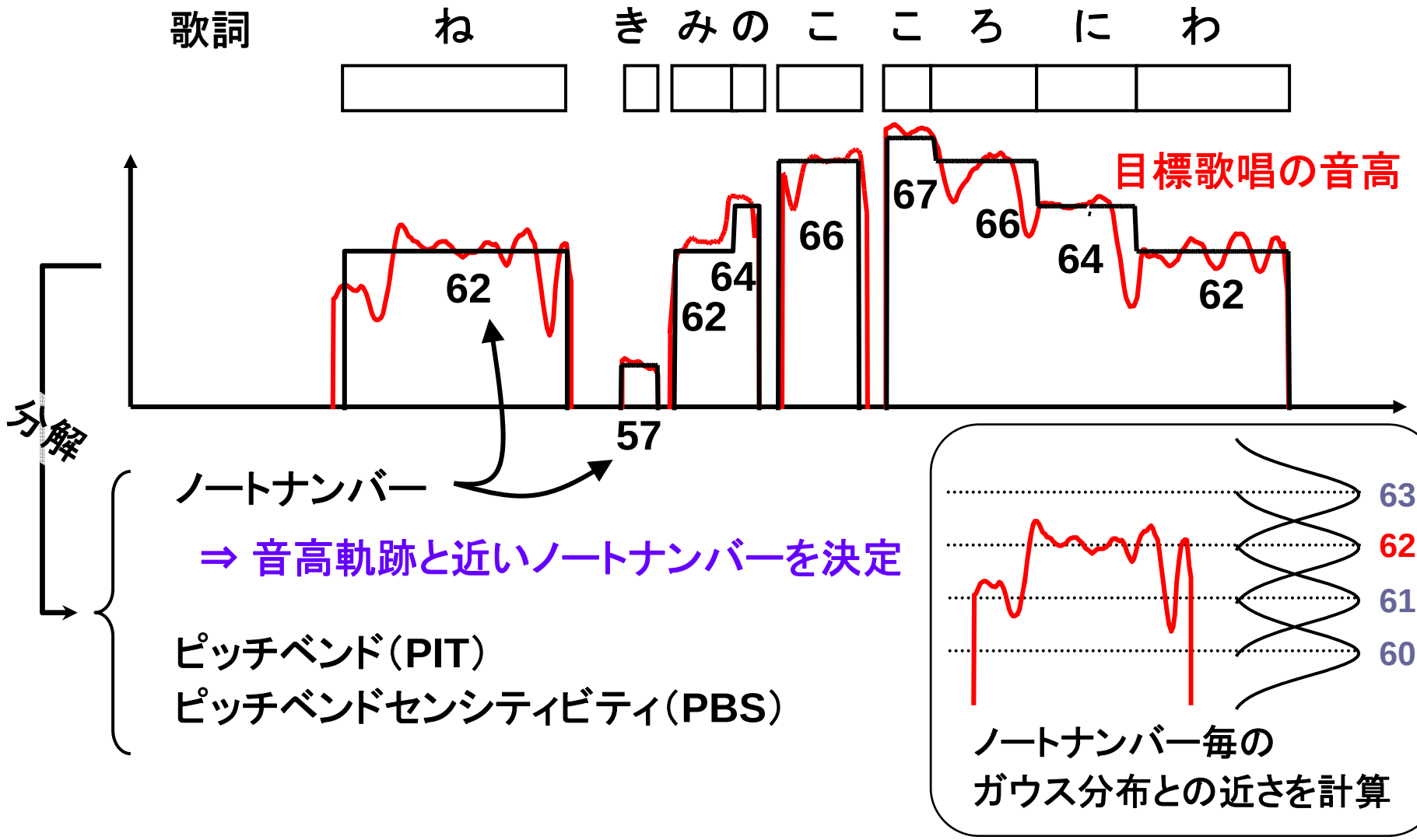
ノートナンバー

⇒ 音節（ひらがな）毎にノートナンバーを決定

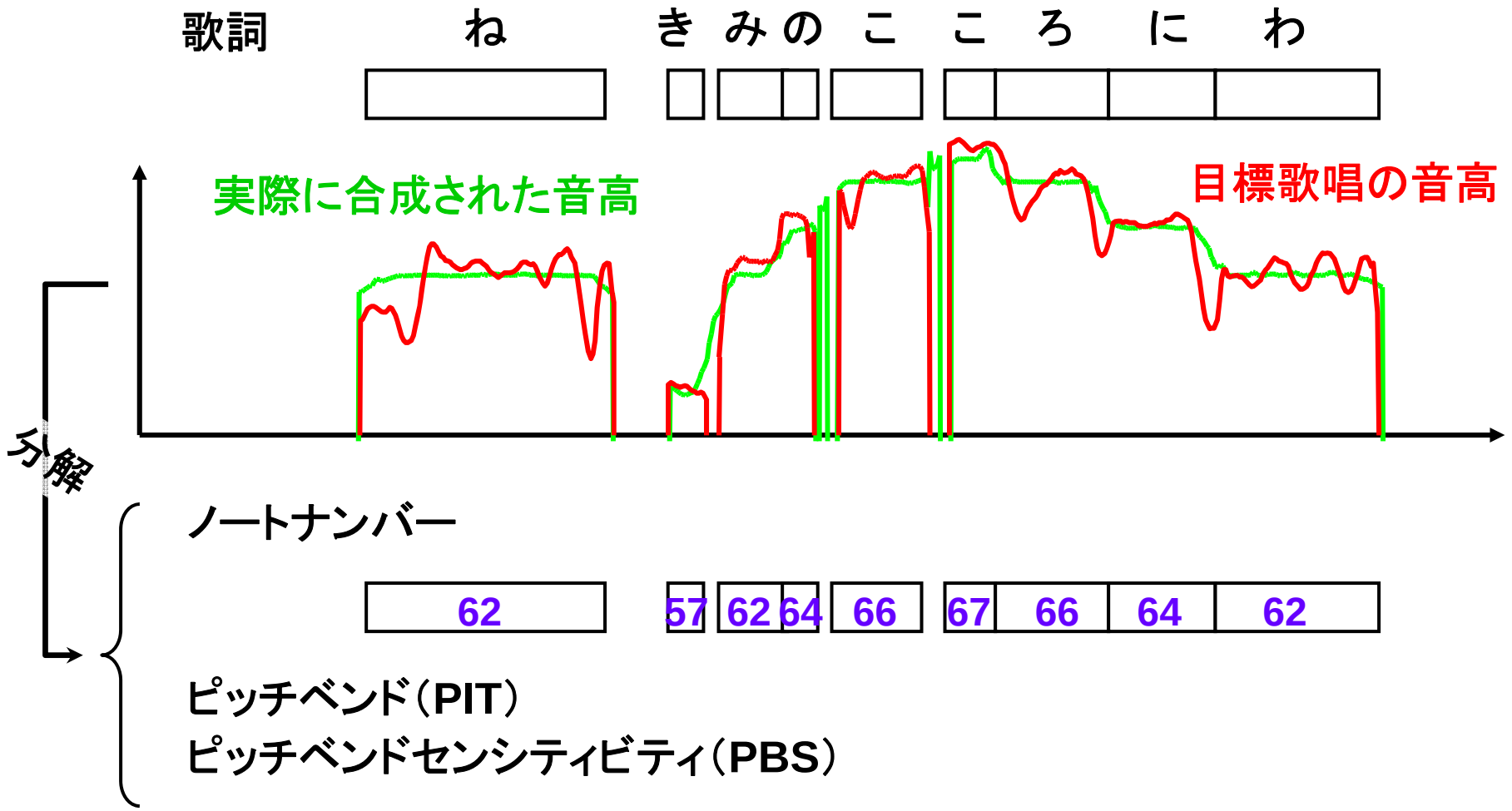
ピッチベンド（PIT）

ピッチベンドセンシティビティ（PBS）

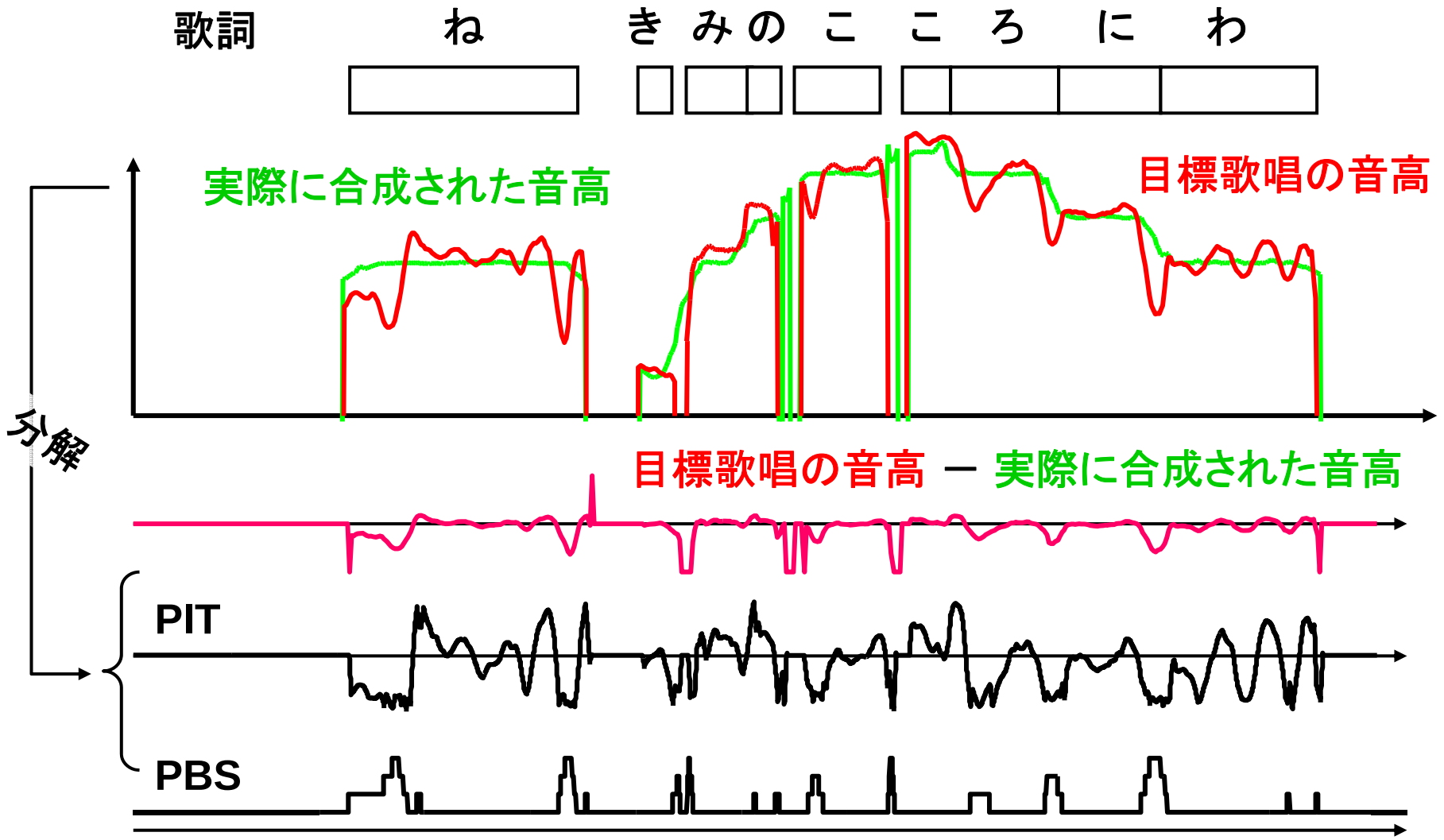
ノートナンバーの決定


歌詞
ね
き
み
の
こ
こ
ろ
に
わ
目標歌唱の音高
62
64
62
66
67
66
64
62
57
分解
ノートナンバー
⇒ 音高軌跡と近いノートナンバーを決定
ピッチベンド（PIT）
ピッチベンドセンシティビティ（PBS）
63
62
61
60
ノートナンバー毎の
ガウス分布との近さを計算

ノートナンバーだけ与えて歌声合成



PIT, PBS の決定


歌詞
ね
き
み
の
こ
こ
ろ
に
わ
実際に合成された音高
目標歌唱の音高
分解
目標歌唱の音高 － 実際に合成された音高
PIT
PBS

歌声合成パラメータ(音量パラメータ)

歌声合成パラメータの自動推定

歌声合成システム

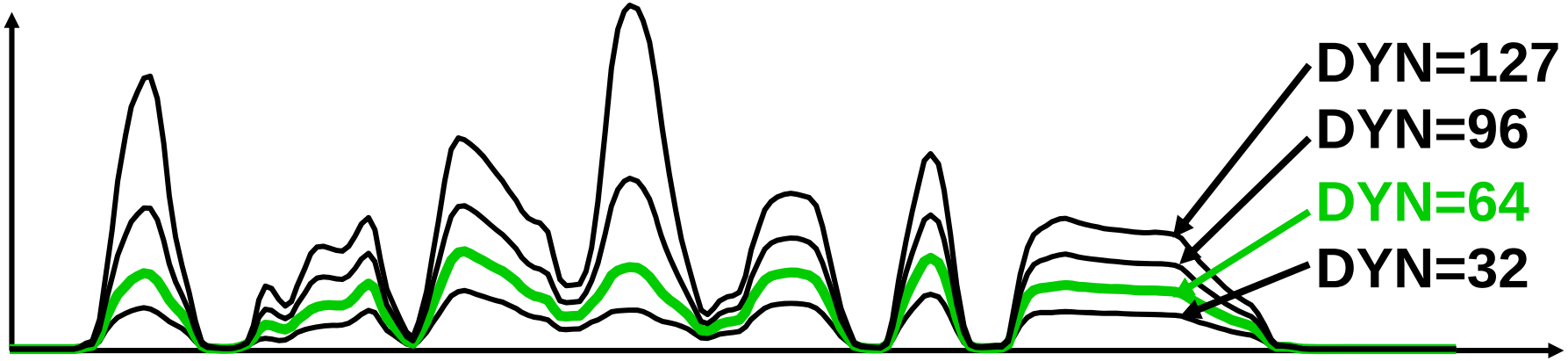
音量パラメータ:Vocaloid2 の場合

ダイナミクス(DYN):0～127 の値

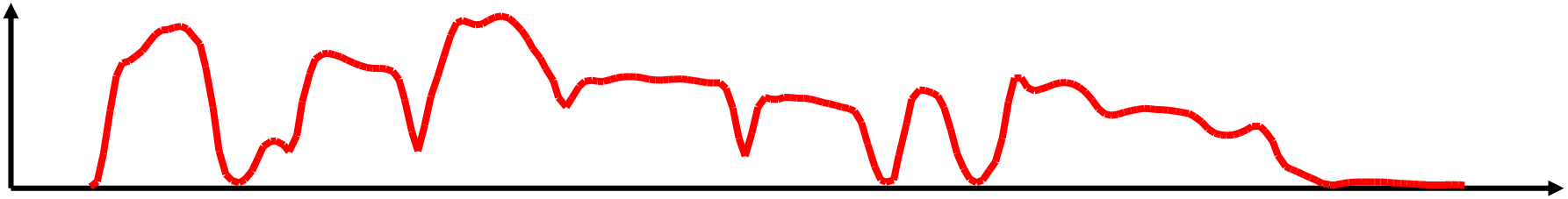
実際に合成して表現範囲を確認

歌詞　た　ち　ど　ま　る　と　き

DYNを0～127まで与えて合成



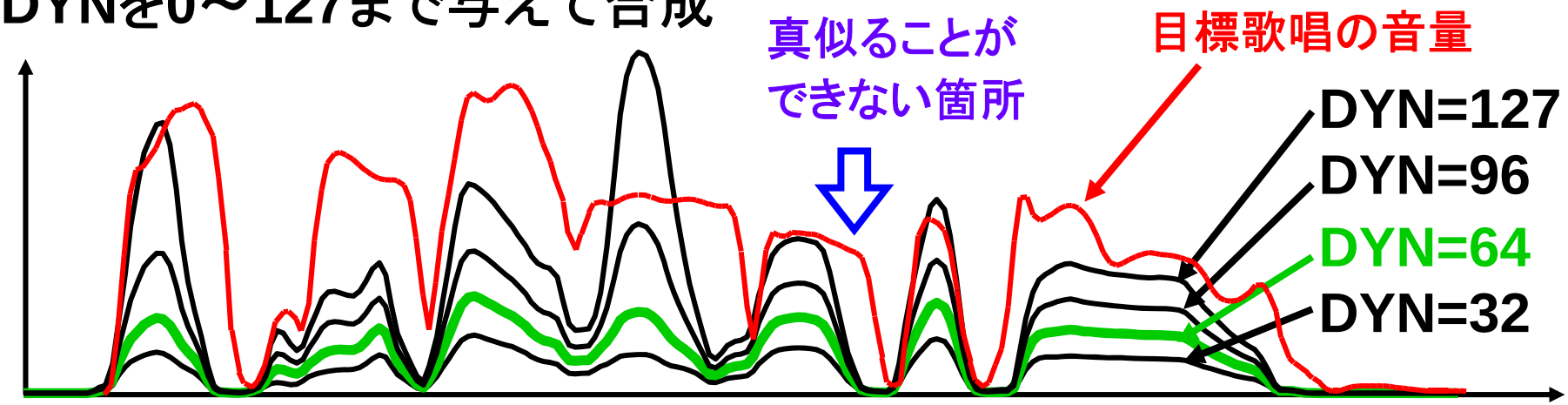
目標歌唱の音量



音量パラメータ推定における問題点

歌詞　た　ち　ど　ま　る　と　き

DYNを0～127まで与えて合成

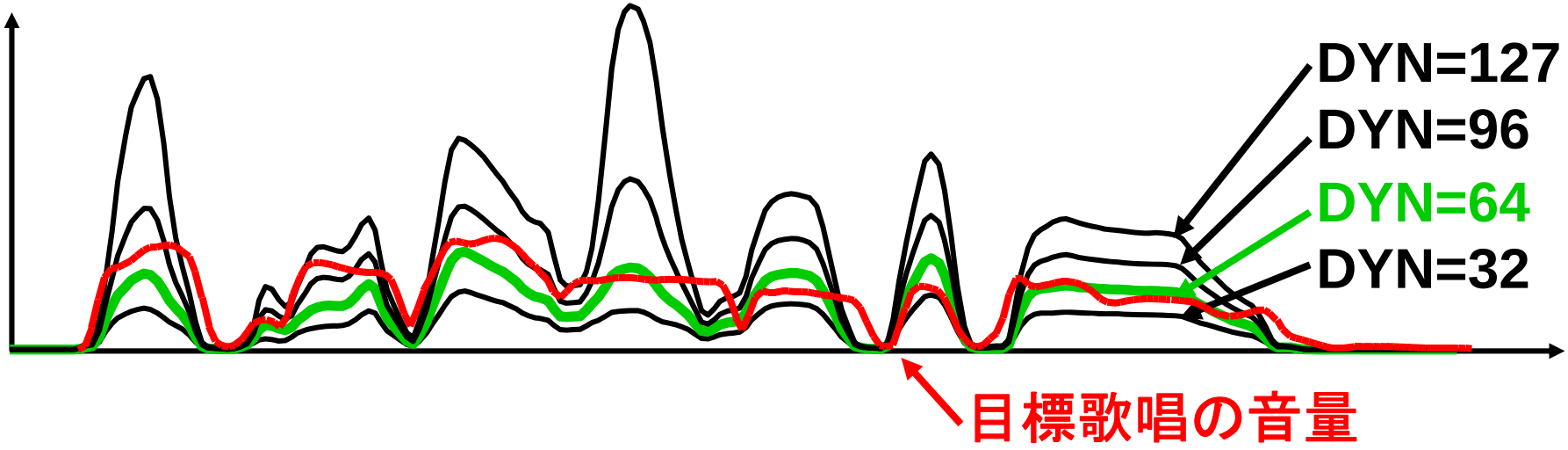


表現範囲に限界がある

⇒目標歌唱の音量を完全に真似ることはできない

本研究における解決法

DYNを0～127まで与えて合成

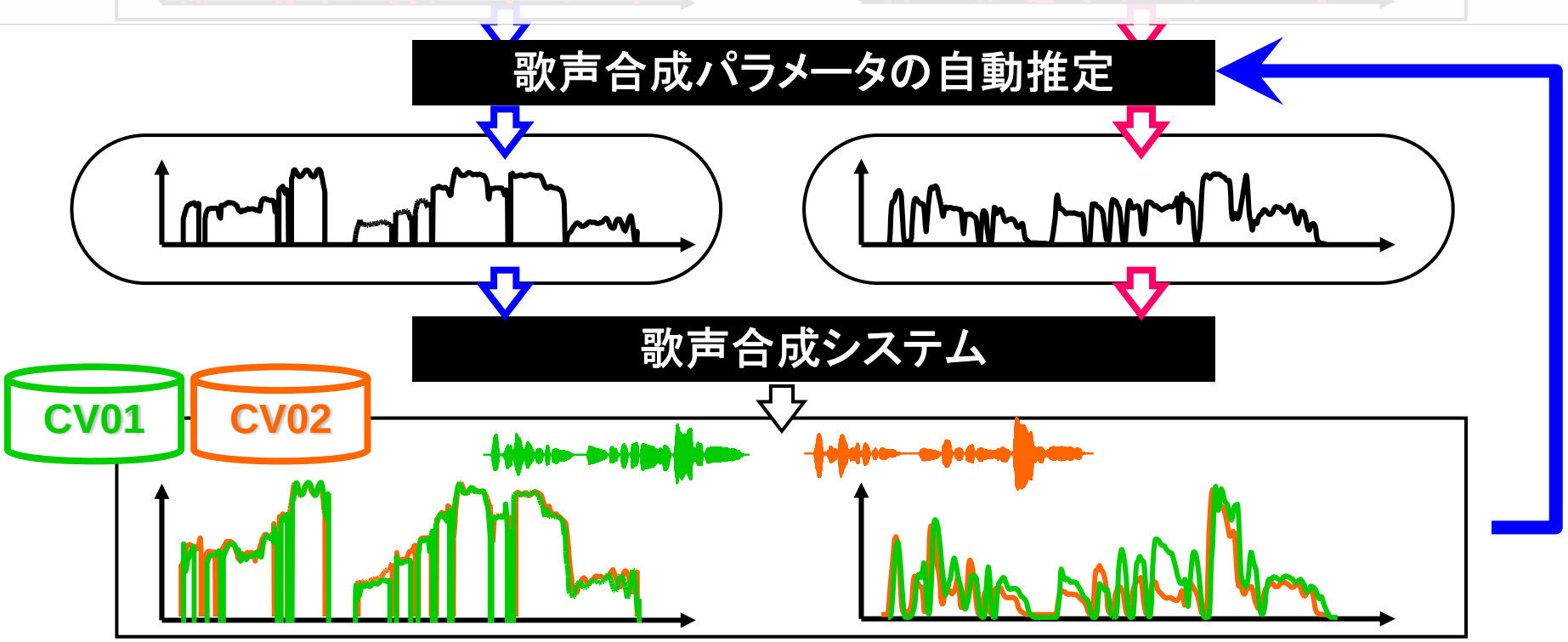


目標歌唱の音量の曲線と
DYNの中心値（=64）の曲線との距離を最小化

⇨ 全体としての再現度を高く

合成パラメータの反復推定

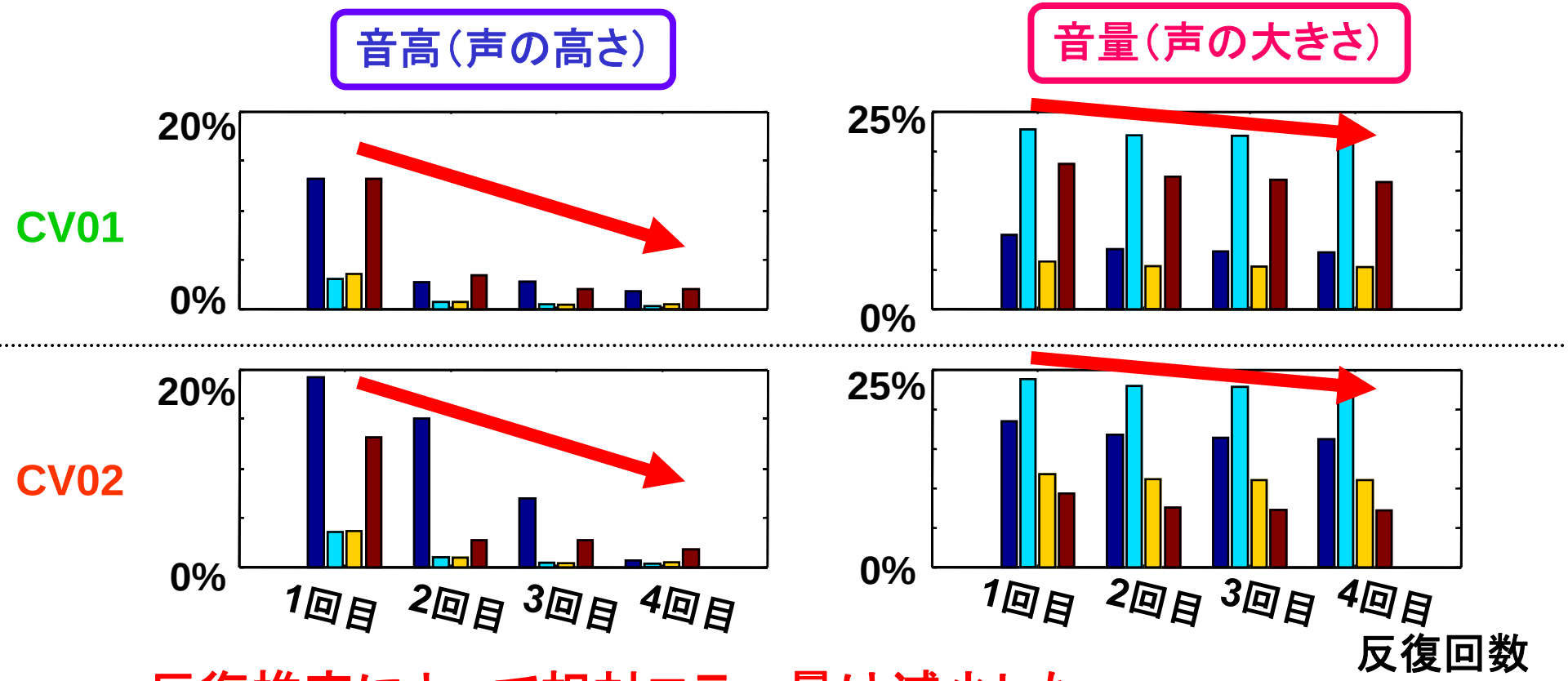
目標歌唱と合成歌唱との
音高・音量の差をパラメータに反映



実験1：反復の効果を確認

実験結果:反復推定による相対エラー量の減少

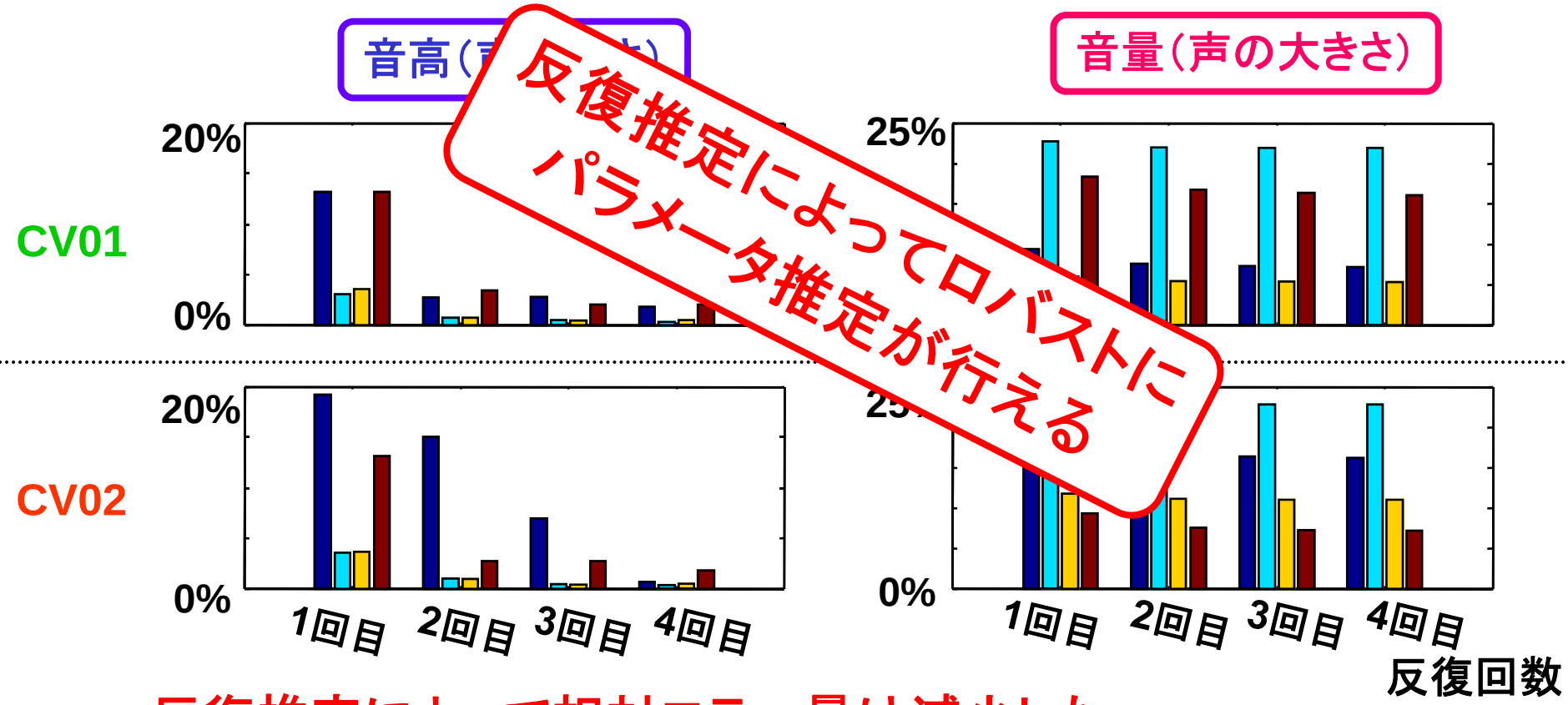
目標歌唱:RWC研究用音楽DB（ポピュラー音楽）
4曲（No.007, No.016, No.054, No.055）冒頭



反復推定によって相対エラー量は減少した

実験結果:反復推定による相対エラー量の減少

目標歌唱:RWC研究用音楽DB（ポピュラー音楽）
4曲（No.007, No.016, No.054, No.055）冒頭



反復推定によって相対エラー量は減少した

本研究の三つのポイント

合成パラメータの反復推定

[redacted]

[redacted]

本研究の三つのポイント

合成パラメータの反復推定

目標歌唱と歌詞の時間的対応付け

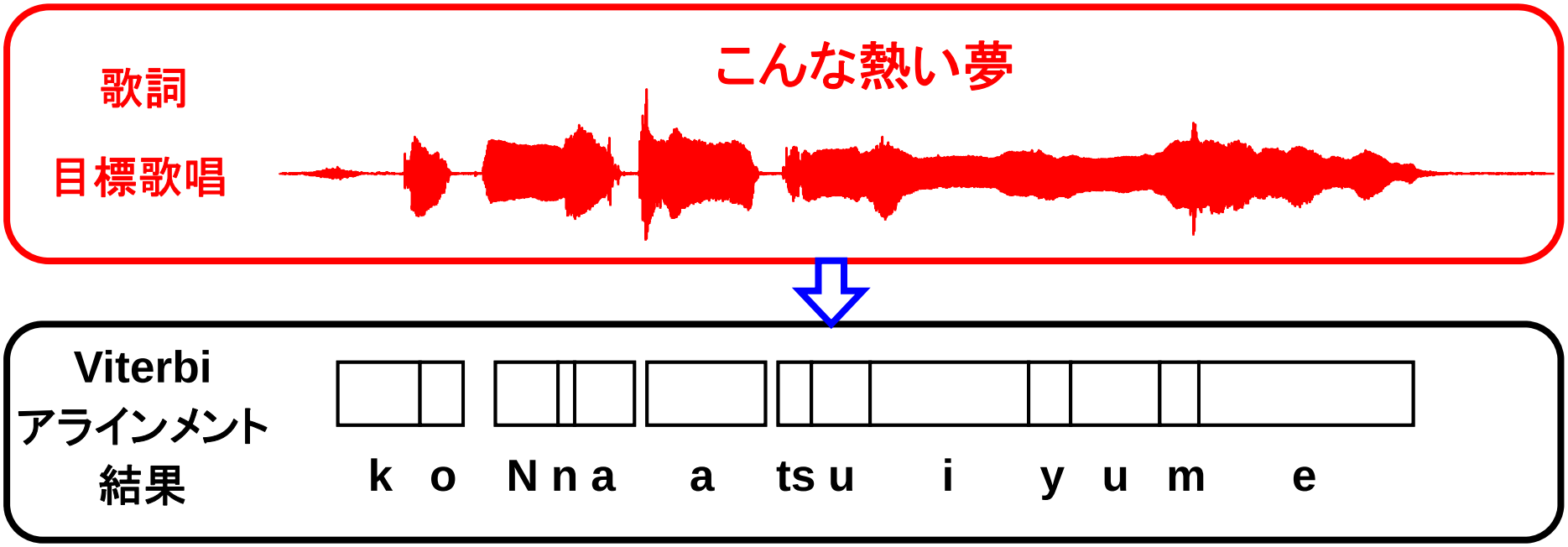
目標歌唱と歌詞の時間的対応付け

- すべての音節の境界を人間が手作業で指定
  - これまでの説明では各音節(「ひらがな」に対応)の始端と終端が決まっていた
  - 手作業で与えるのは大変

⇩

歌詞さえ与えれば音節の境界を自動推定

処理の流れ：Viterbi アラインメント (HMM)



※フレーズを超えるなどの大きな誤りは手作業で修正

処理の流れ：母音の始端と終端を利用して合成

歌詞　こんな熱い夢

目標歌唱

Viterbi
アラインメント
結果

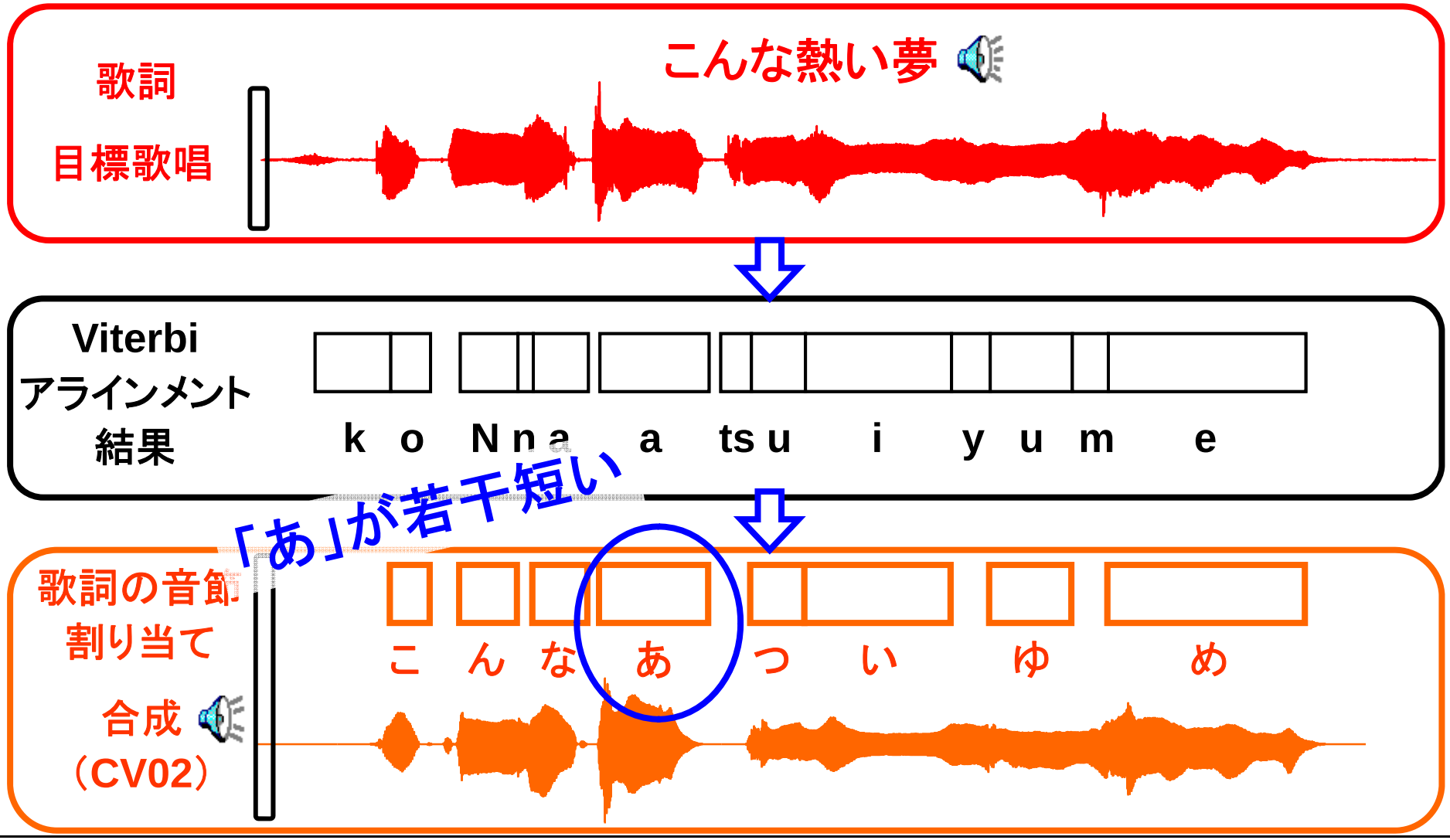
k o N n a a ts u i y u m e

歌詞の音節
割り当て

こ ん な あ つ い ゆ め

合成
（CV02）

問題点


歌詞
こんな熱い夢
目標歌唱
Viterbi
アラインメント
結果
k o N n a a ts u i y u m e
「あ」が若干短い
歌詞の音節
割り当て
こ ん な あ つ い ゆ め
合成
（CV02）

原因

歌詞

こんな熱い夢

目標歌唱

Viterbi
アラインメント
結果

k o N n a a

Viterbi アラインメントの性能
合成システムの特性

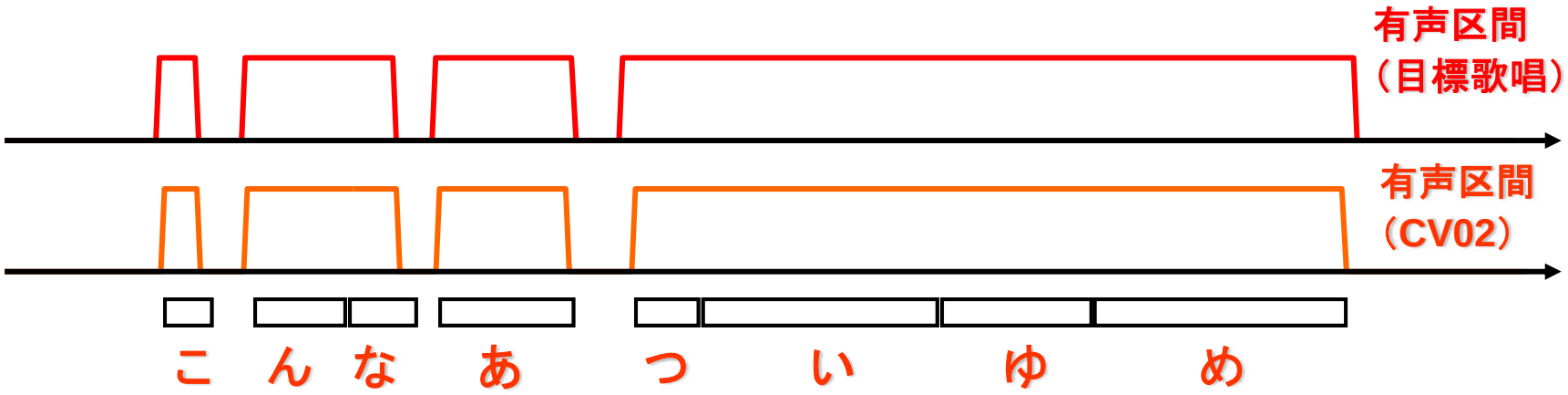
「あ」が若干短い

歌詞の音韻
割り当て

こ ん な あ つ い ゆ め

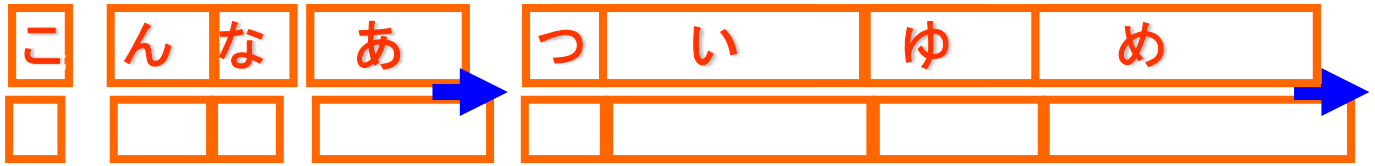
合成
（CV02）

本研究の解決法：有声区間のずれを補正



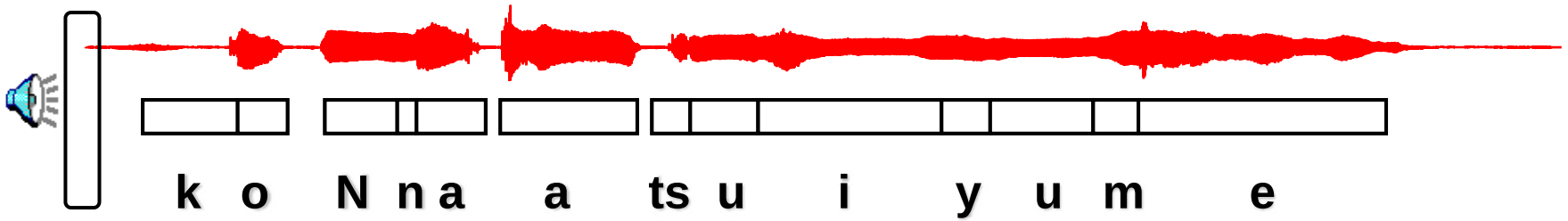
1. 有声区間中は前後の二音節を接続

2. 目標と合成の有声区間が一致するように各音節の始端と終端を伸縮

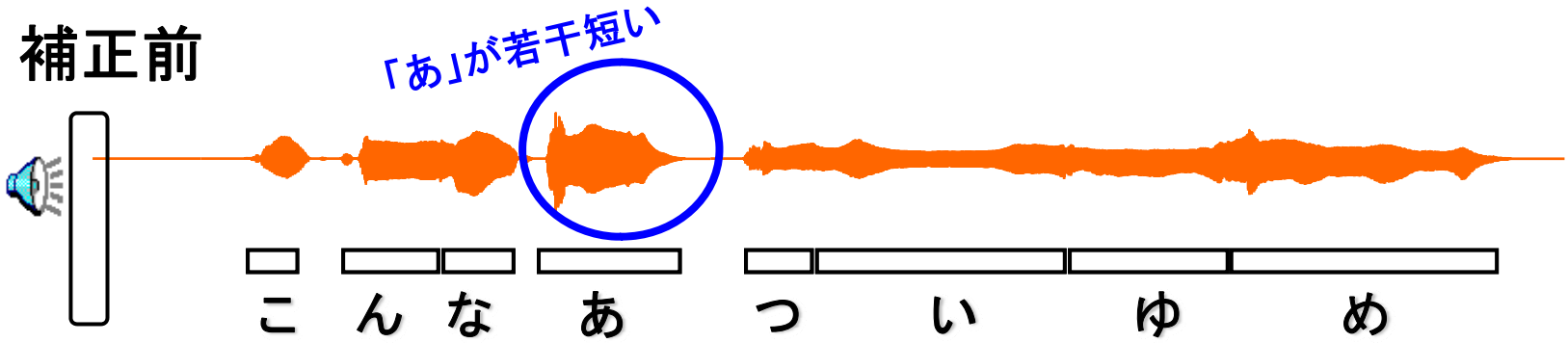


有声区間のずれを補正した結果

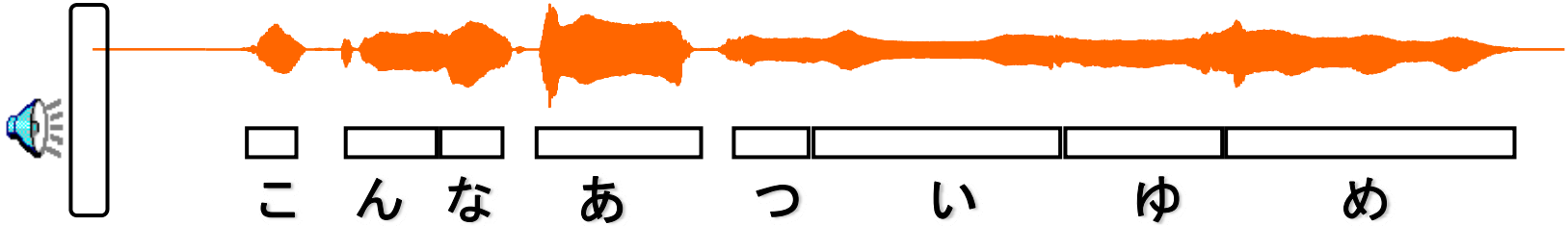
目標歌唱



補正前

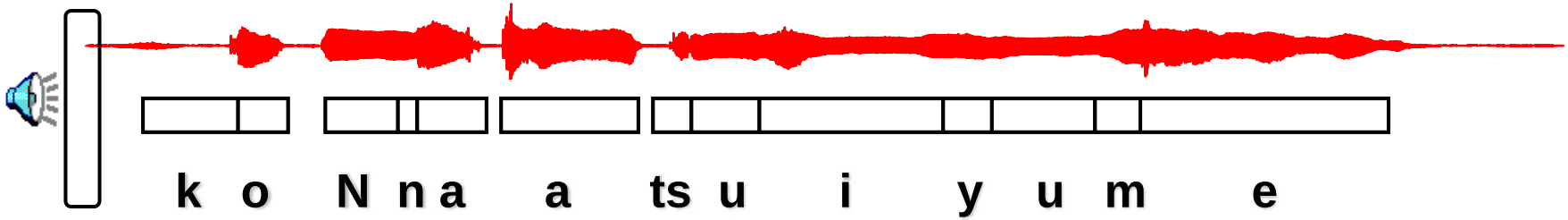


補正後



有声区間のずれを補正した結果

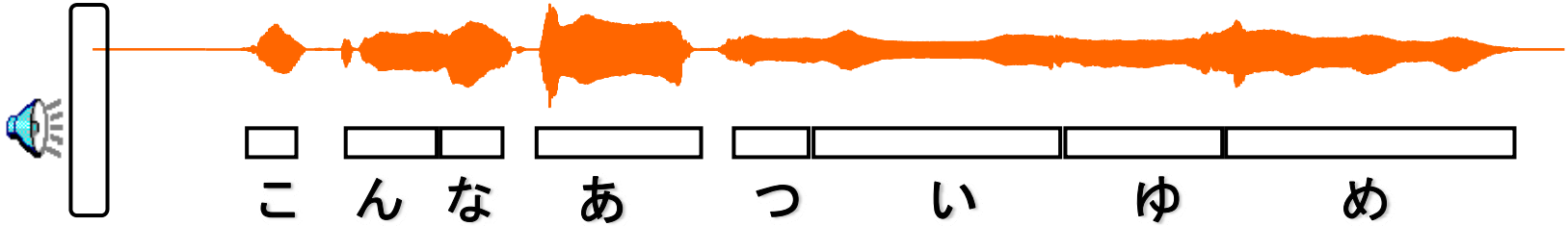
目標歌唱



補正前

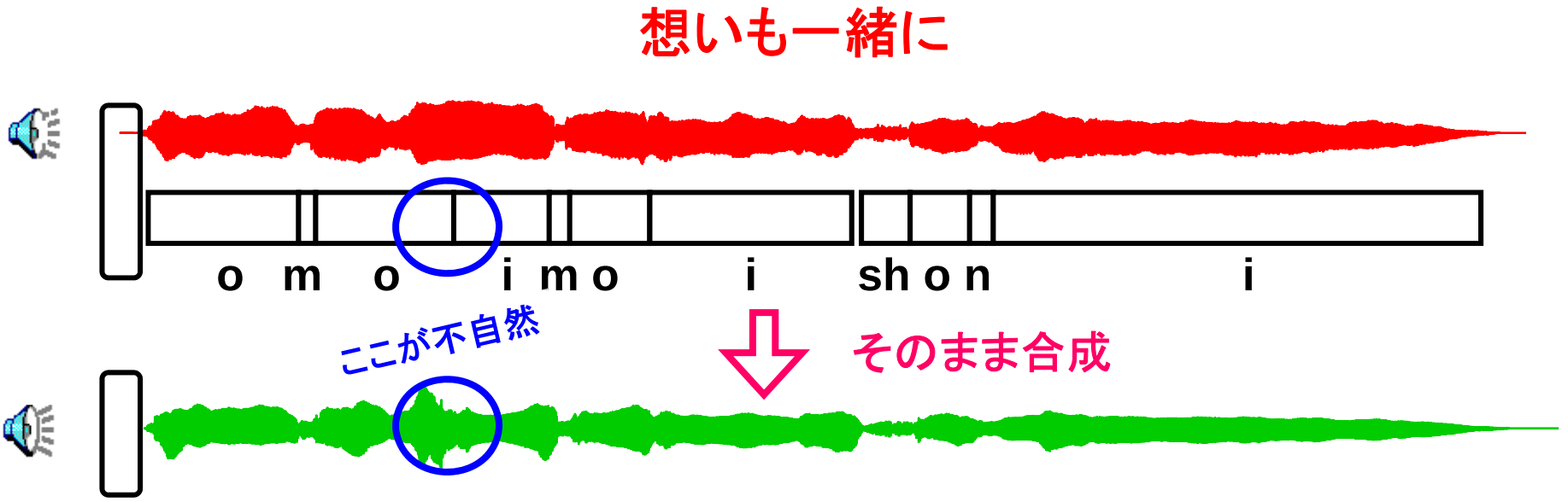
有声区間のずれがなくなり
品質がより向上

補正後



もう一つの問題

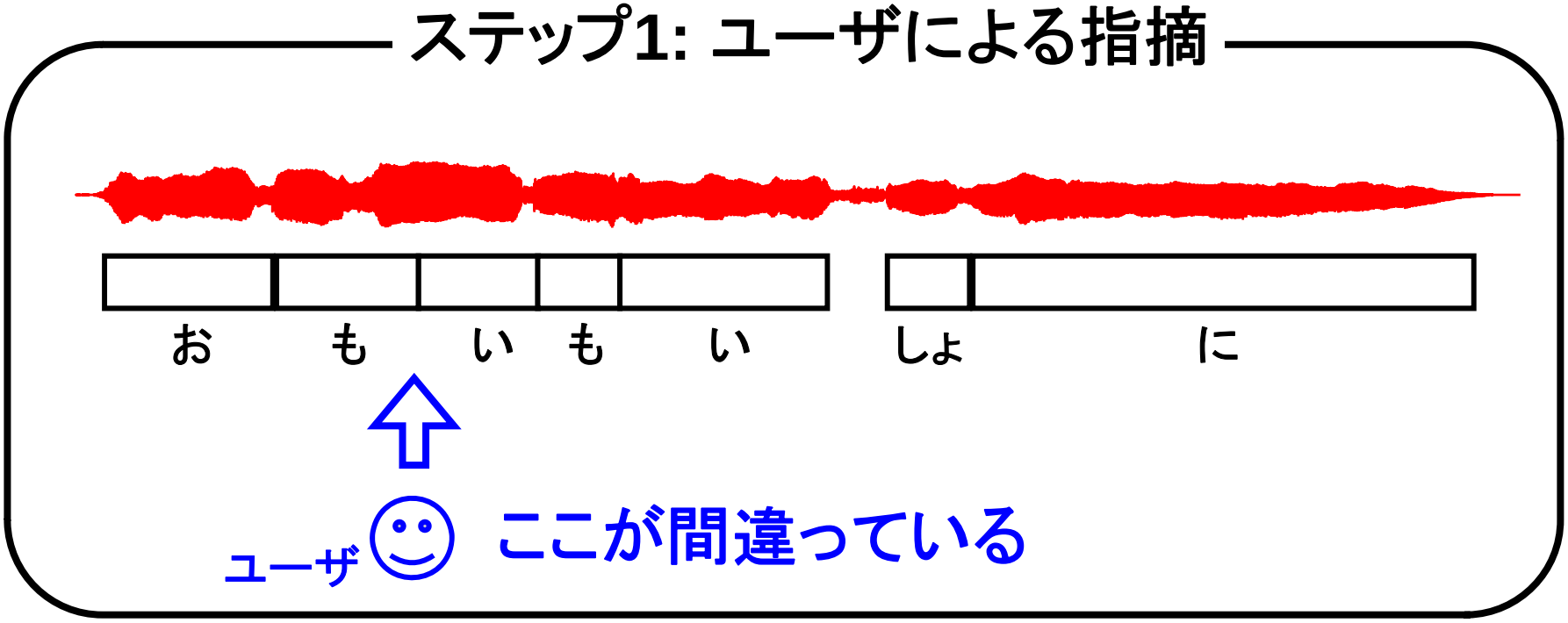
- Viterbiアラインメント結果に誤りが生じる



従来、自動推定結果の誤りへの
対処は考えられていなかった

本研究の解決法: 音節境界の誤り訂正

- ユーザが誤り箇所を指摘する
- 新しい境界候補を自動的に推定して再提示

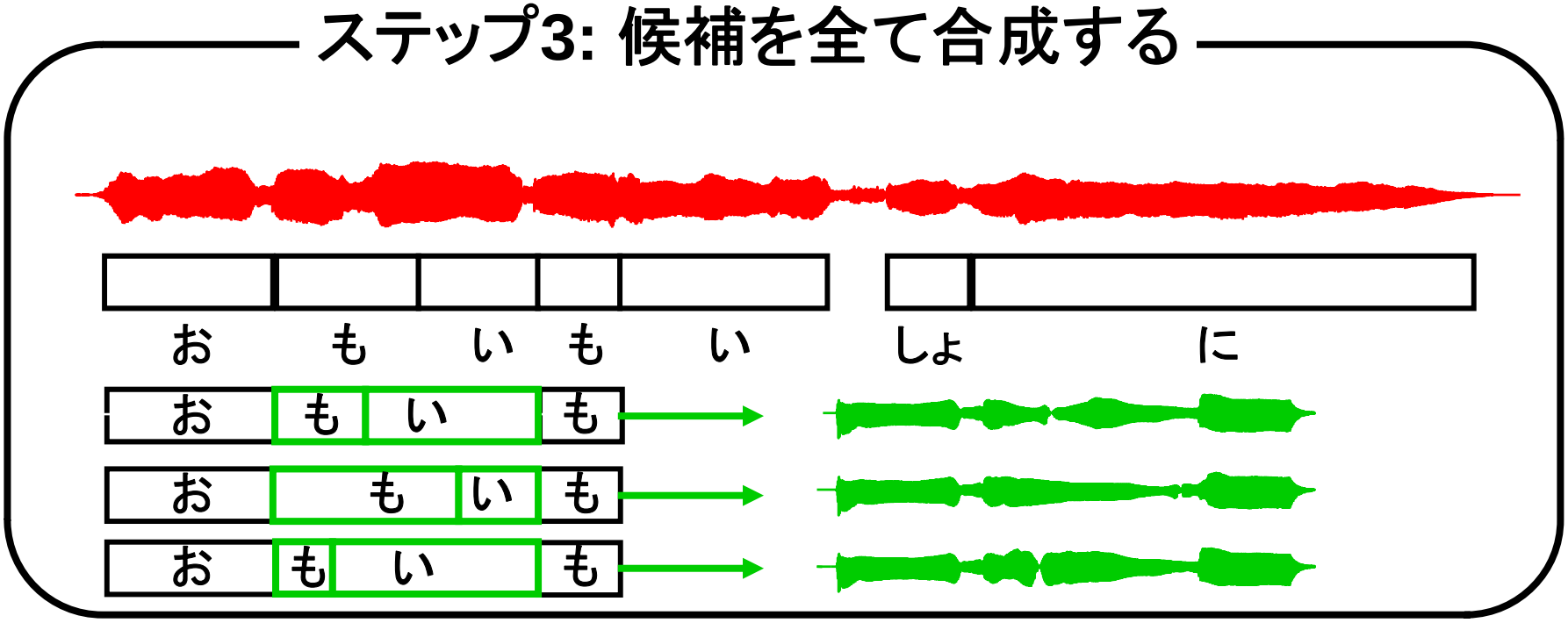


本研究の解決法: 音節境界の誤り訂正

- ユーザが誤り箇所を指摘する
- 新しい境界候補を自動的に推定して再提示

本研究の解決法: 音節境界の誤り訂正

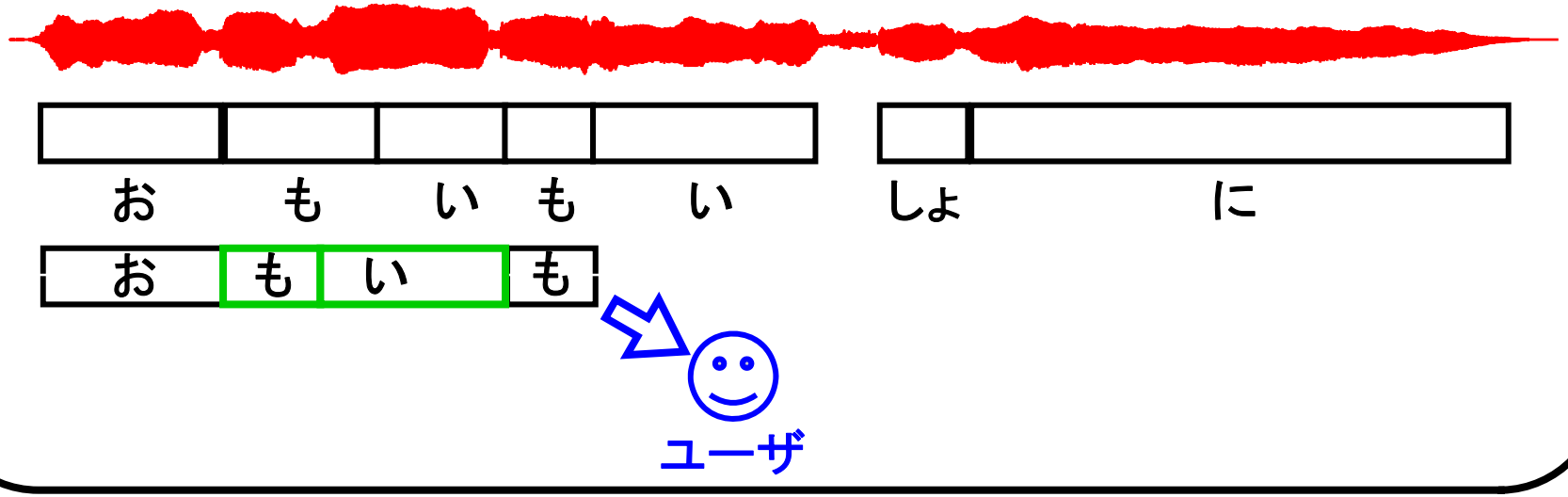
- ユーザが誤り箇所を指摘する
- 新しい境界候補を自動的に推定して再提示



本研究の解決法: 音節境界の誤り訂正

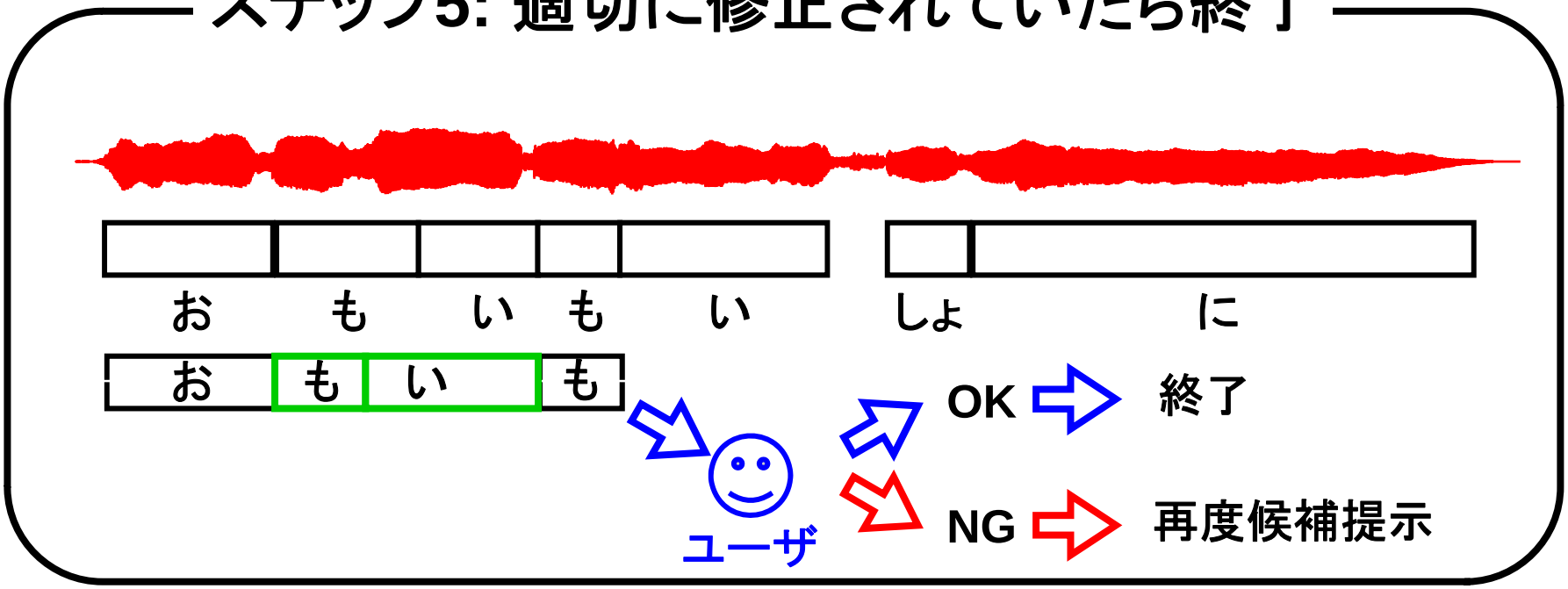
- ユーザが誤り箇所を指摘する
- 新しい境界候補を自動的に推定して再提示

ステップ4: 距離最小の候補をユーザに提示



本研究の解決法: 音節境界の誤り訂正

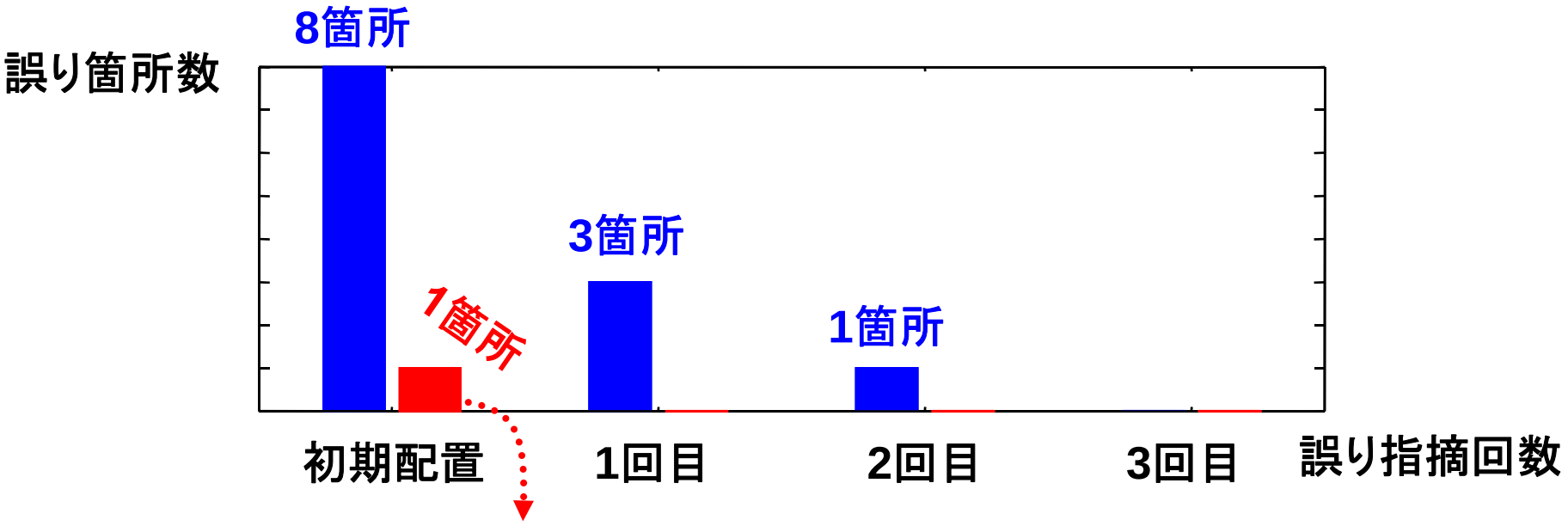
- ユーザが誤り箇所を指摘する
- 新しい境界候補を自動的に推定して再提示



実験2：音節境界の誤り訂正回数

実験結果：音節境界誤り訂正における指摘回数

- 目標歌唱：RWC研究用音楽DB（ポピュラー音楽）
  No.007, No.016（歌詞の一番）
  全166音節　全128音節

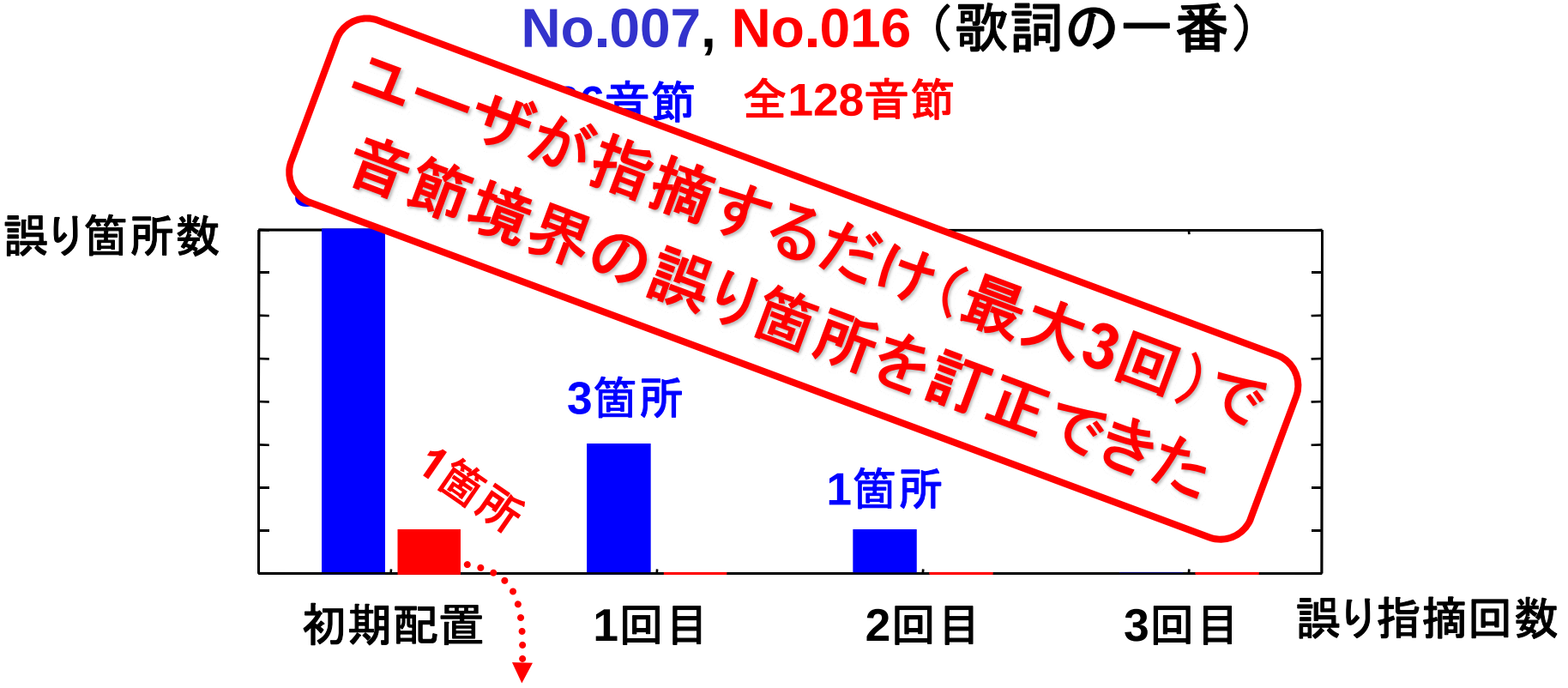


フレーズを超えるなどの大きな誤り二箇所を手作業で修正

実験結果：音節境界誤り訂正における指摘回数

- 目標歌唱：RWC研究用音楽DB（ポピュラー音楽）

No.007, No.016（歌詞の一番）



本研究の三つのポイント

合成パラメータの反復推定

目標歌唱と歌詞の時間的対応付け

[redacted]

本研究の三つのポイント

合成パラメータの反復推定

目標歌唱と歌詞の時間的対応付け

歌唱力補正

歌唱力補正機能

- 歌唱力が高くないユーザでも使えるように
- 自分とは違うスタイルの歌唱を生成できるように

⇩

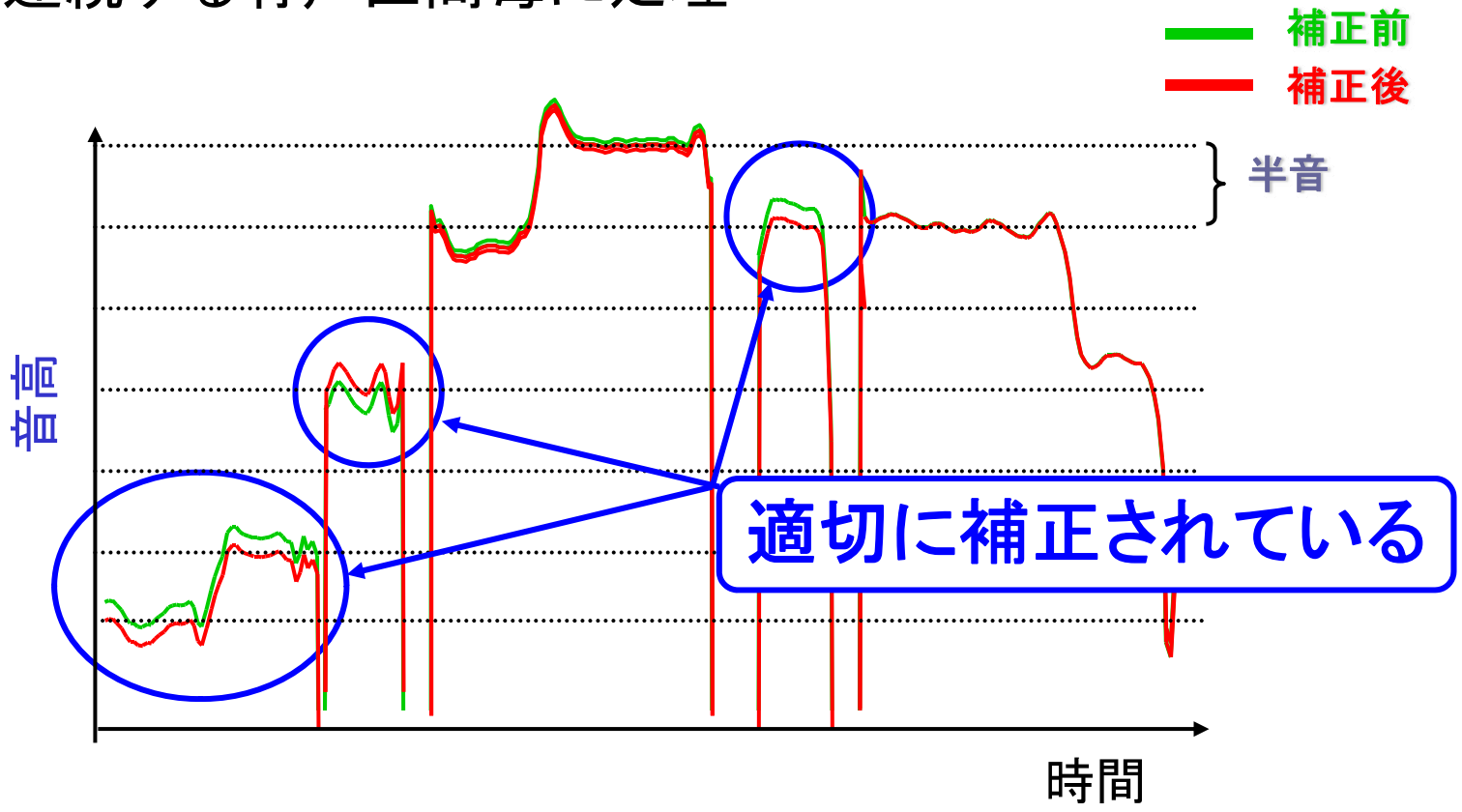
音高変更機能

歌唱スタイル変更機能

音高変更機能

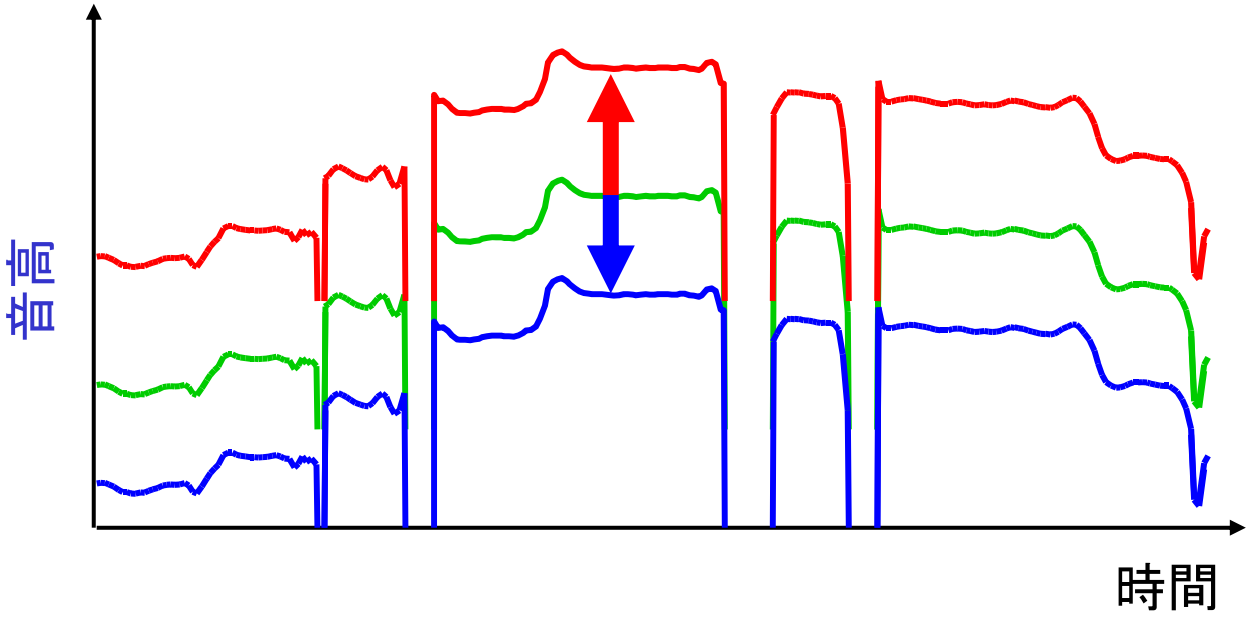
音高変更機能:調子はずれ(off-pitch)の補正

- 音高遷移が半音単位となるように補正
  - 連続する有声区間毎に処理



音高変更機能：音高トランスポーズ

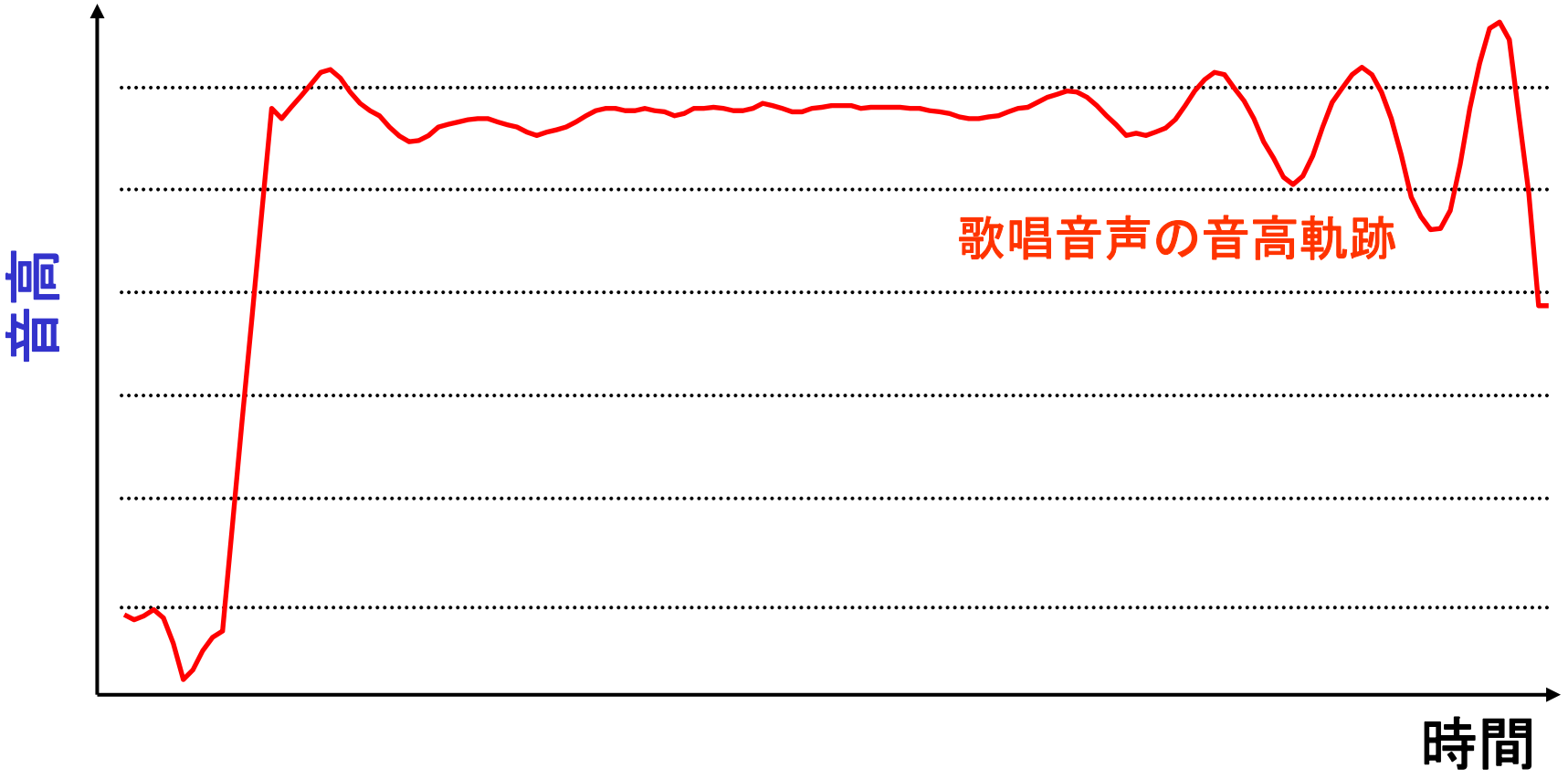
- 全体、もしくはユーザが指定した区間の音高を変更
  - 声域の違いを克服できる



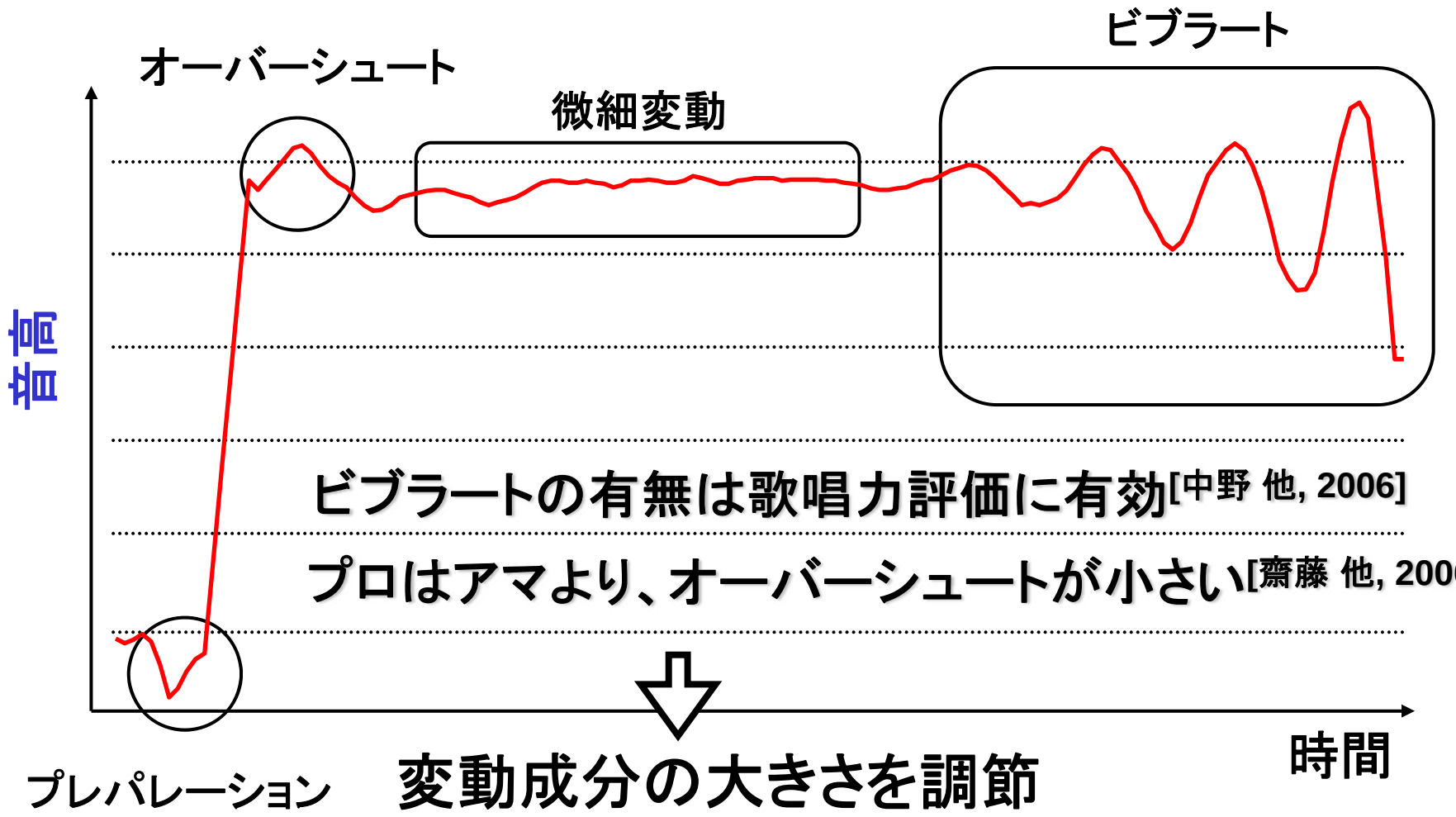
歌唱スタイル変更機能

歌唱スタイルの変更機能

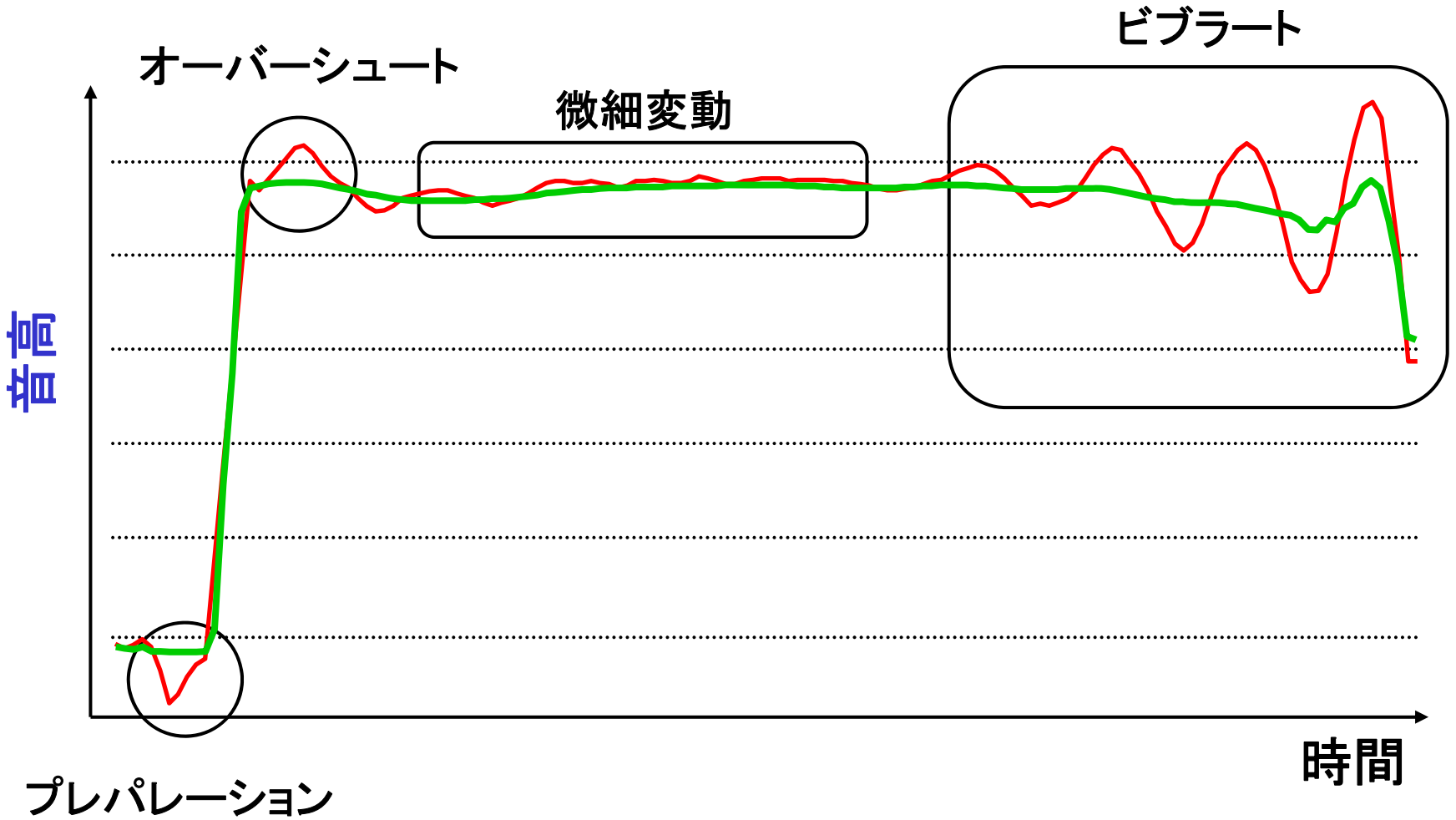
- 音高・音量軌跡を変更することで、歌唱力を補正



歌唱音声の音高における動的変動成分[齋藤 他, 2008]

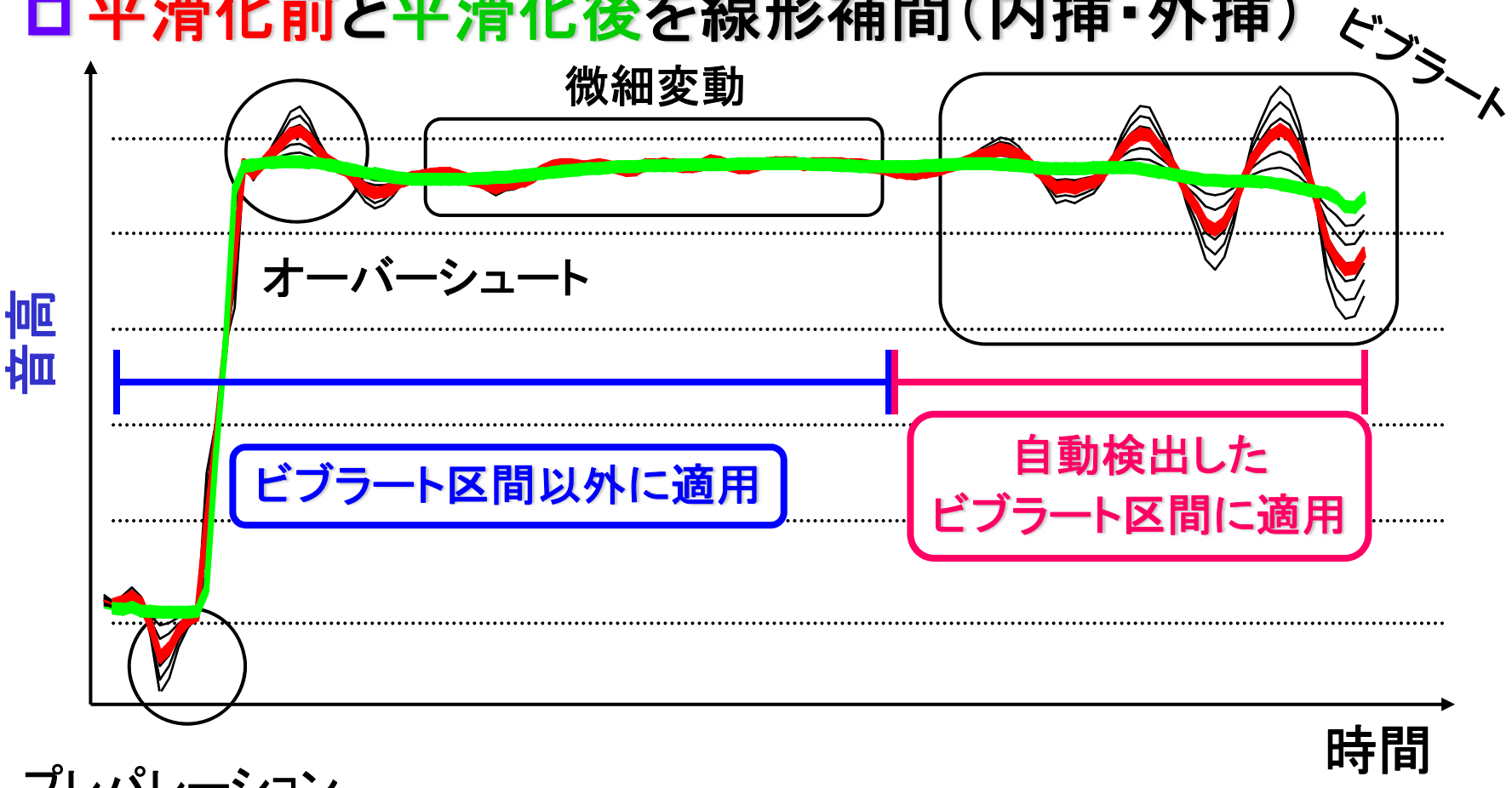


音高軌跡を平滑化



動的変動成分を強調・抑制

□ 平滑化前と平滑化後を線形補間（内挿・外挿）



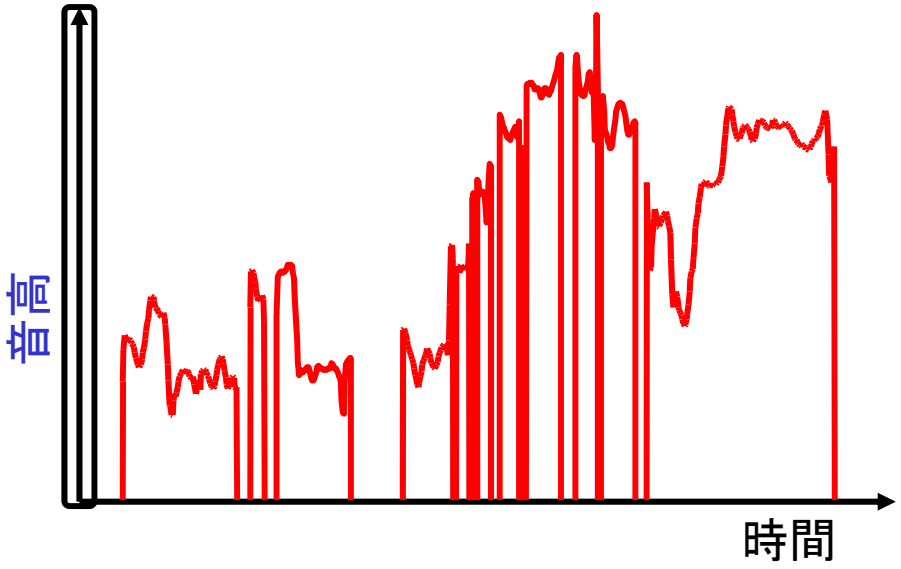
※音量も同様に処理する

デモ：音高・歌唱スタイル変更機能の適用

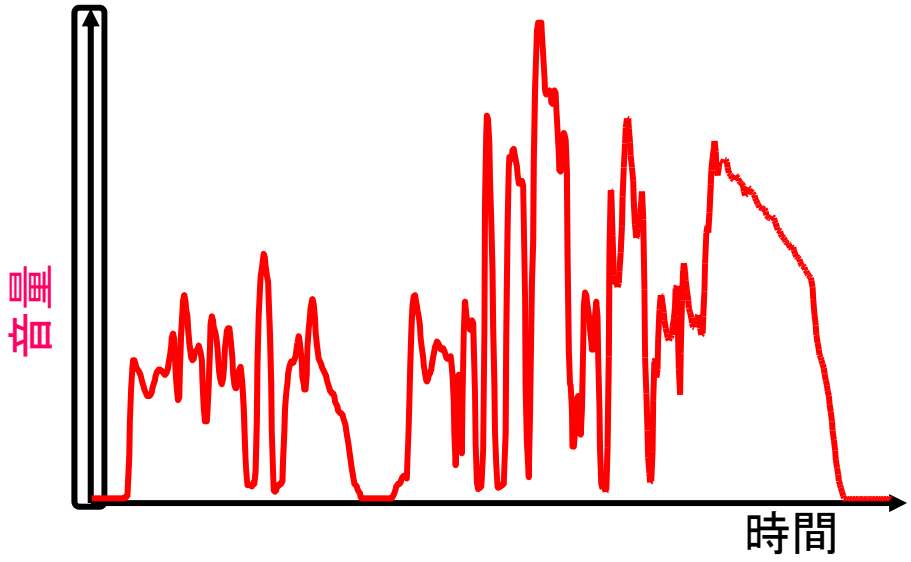
歌詞：今も せつない姿 探しているよ

ユーザ歌唱

音高（声の高さ）



音量（声の大きさ）

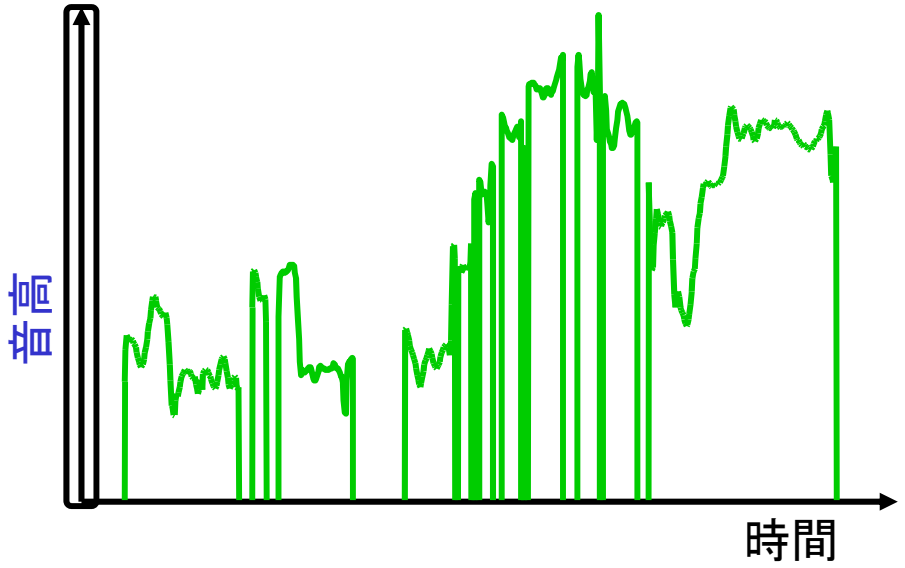


デモ：音高・歌唱スタイル変更機能の適用

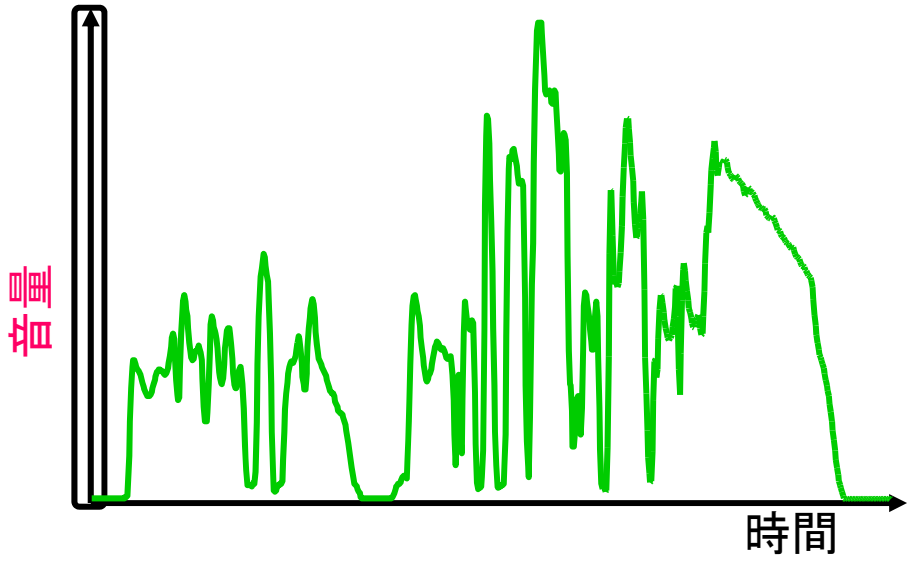
歌詞：今も せつない姿 探しているよ

オクターブ上げて合成（CV01）

音高（声の高さ）



音量（声の大きさ）

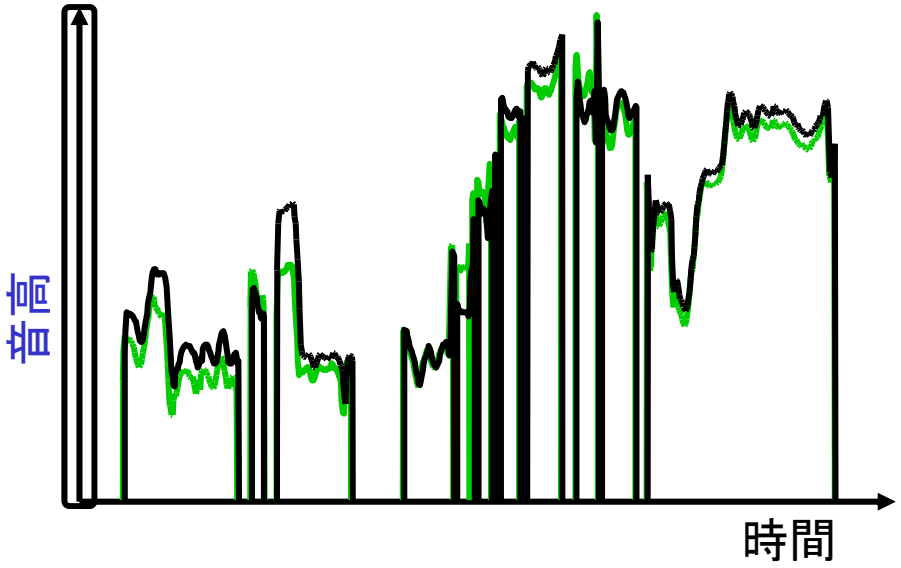


デモ：音高・歌唱スタイル変更機能の適用

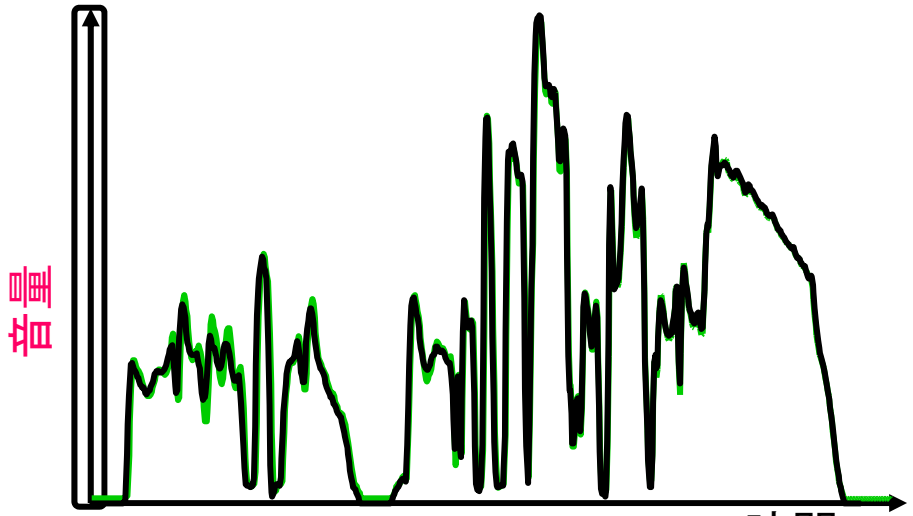
歌詞：今も せつない姿 探しているよ

全ての修正機能を適用して合成（CV01）

音高（声の高さ）



音量（声の大きさ）



本研究の三つのポイント

合成パラメータの反復推定

目標歌唱と歌詞の時間的対応付け

歌唱力補正

今後の展望

- **歌声研究の基本ツールとしての VocaListener**
  - 心理実験用の刺激生成
    - 歌唱の個人性知覚の秘密を探る
    - うまい歌唱の秘密を探る（歌唱力評価）

- **歌声合成の支援ツールとしての VocaListener**
  - メタ歌声合成システムの実現
  - より人間らしい合成歌唱の実現
    - ブレス自動検出法によるブレス付与
  - 歌唱力補正機能の充実

VocaListener

ユーザ歌唱を真似る歌声合成パラメータを自動推定するシステムの提案

中野倫靖, 後藤真孝
（産業技術総合研究所）

2008年5月28日
第75回音楽情報科学研究会(SIGMUS)
第128回ヒューマンコンピュータインタラクション研究会 (SIGHCI)