

# 市販楽曲中の**歌声の分離**と**音高推定** に基づく**歌唱表現編集**システム

池宮 由楽 糸山 克寿 吉井 和佳

京都大学 大学院情報学研究科

知能情報学専攻 音声メディア分野



# 池宮由楽君の仕事です

当研究室のエースですが  
この春で就職します

諸事情で参加できませんので  
指導教員が代理発表いたします







**基盤技術**  
**音楽CDの歌声分離・音高推定**

**アプリケーション**  
**プロ歌手の歌声を編集**

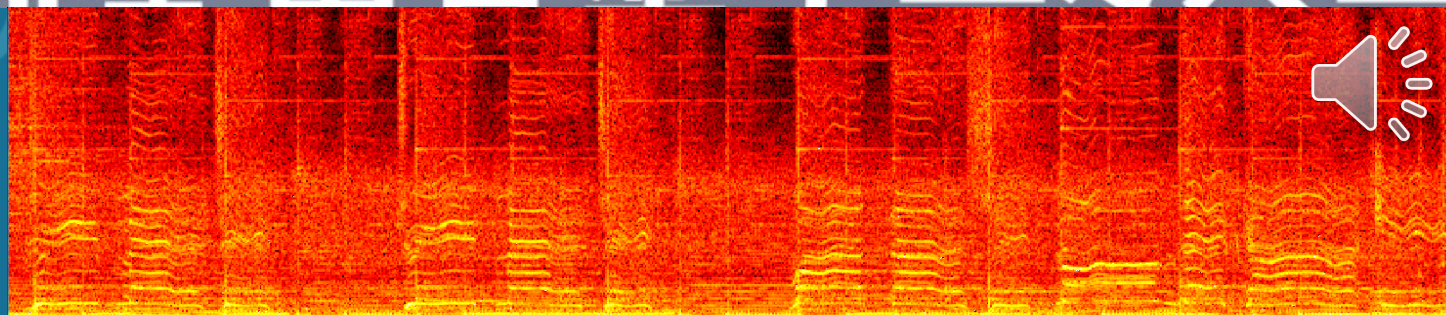


English version is [here](#).

# RWC-MDB-2001 No.33

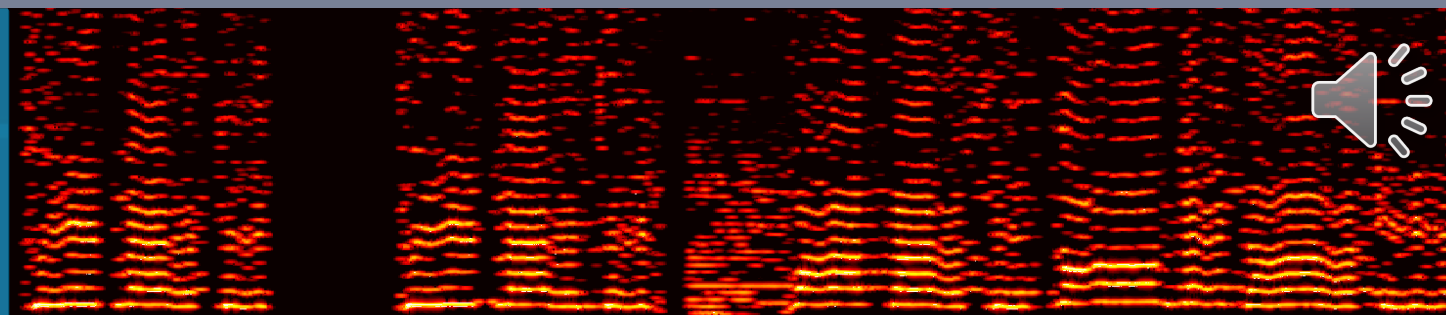
## RWC研究用音源データベース

オリジナル

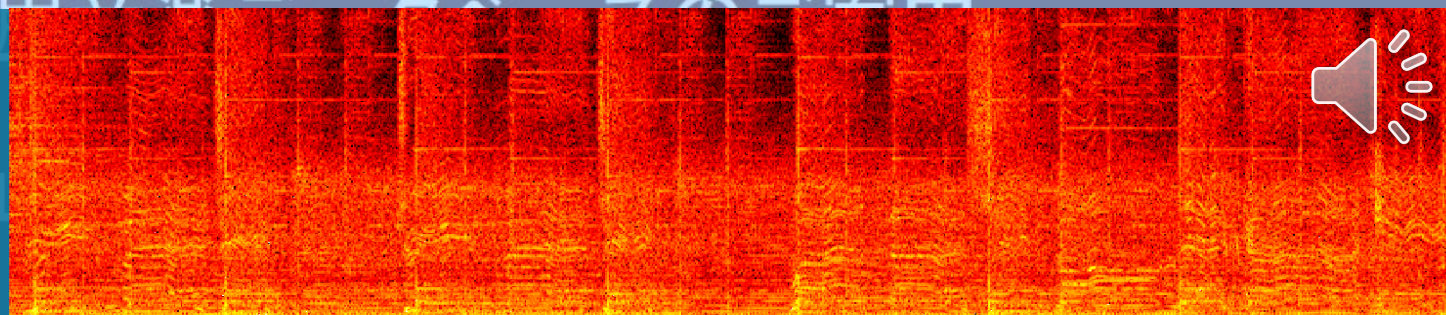


歌声

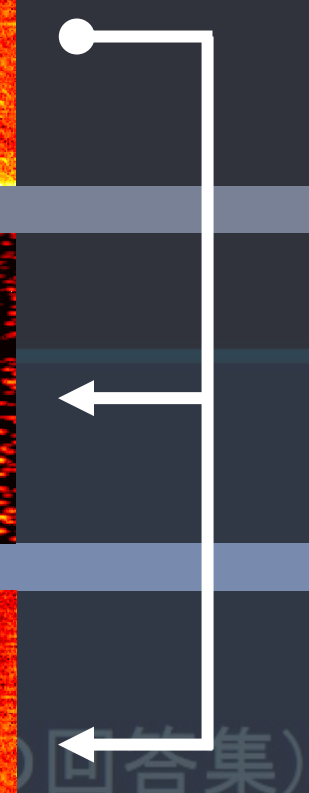
Contents:



伴奏



- [RWC研究用音源データベースの概要](#)
- [利用方法](#)
- [FAQ: Frequency](#)
- [文献](#)
- [連絡先](#)
- [メーリングリスト](#)





# ずっと / aiko

オリジナル



歌声



音高



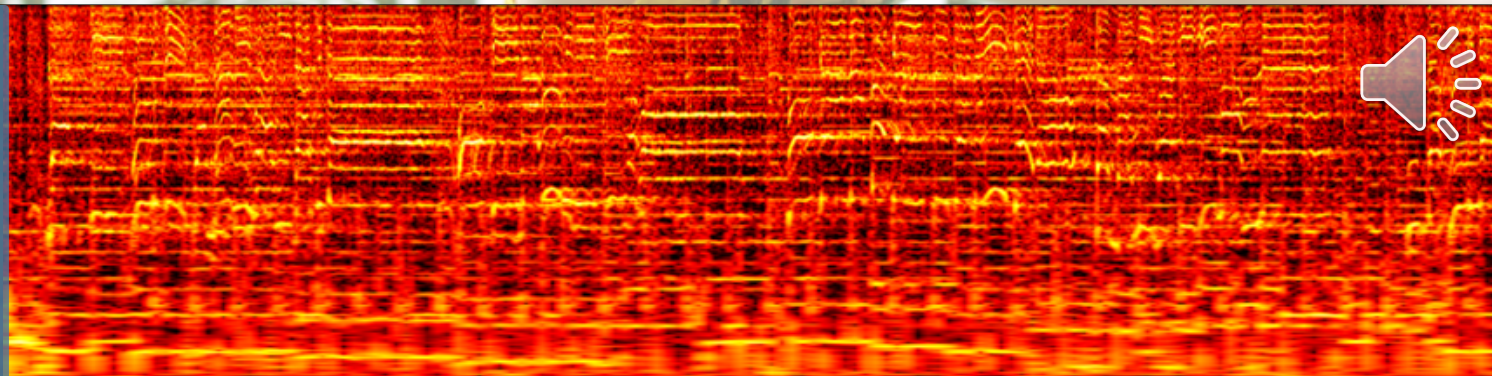
伴奏



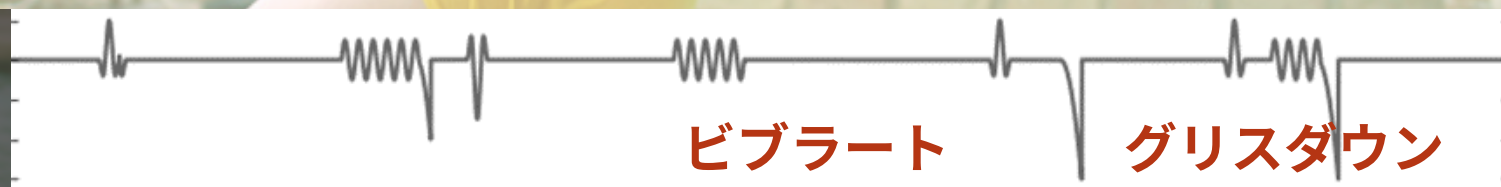


# ラッキープール / JUDY AND MARY

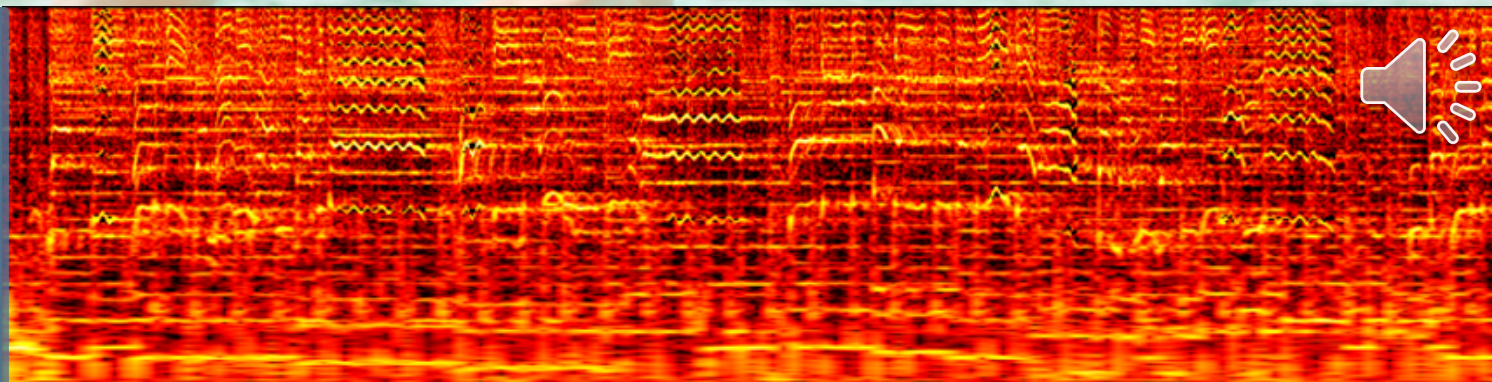
編集前



歌唱表現



編集後





# 能動的音楽鑑賞インタフェース

# Songgle

## 音楽の「理解」

繰り返し区間検出・ビート時刻推定

コード進行推定・歌声音高推定

# Active Music Listening

# Interface Visualizing

## 鑑賞支援

# Music Scene Descriptions

# 能動的音楽鑑賞インタフェース

## 音楽の「編集」

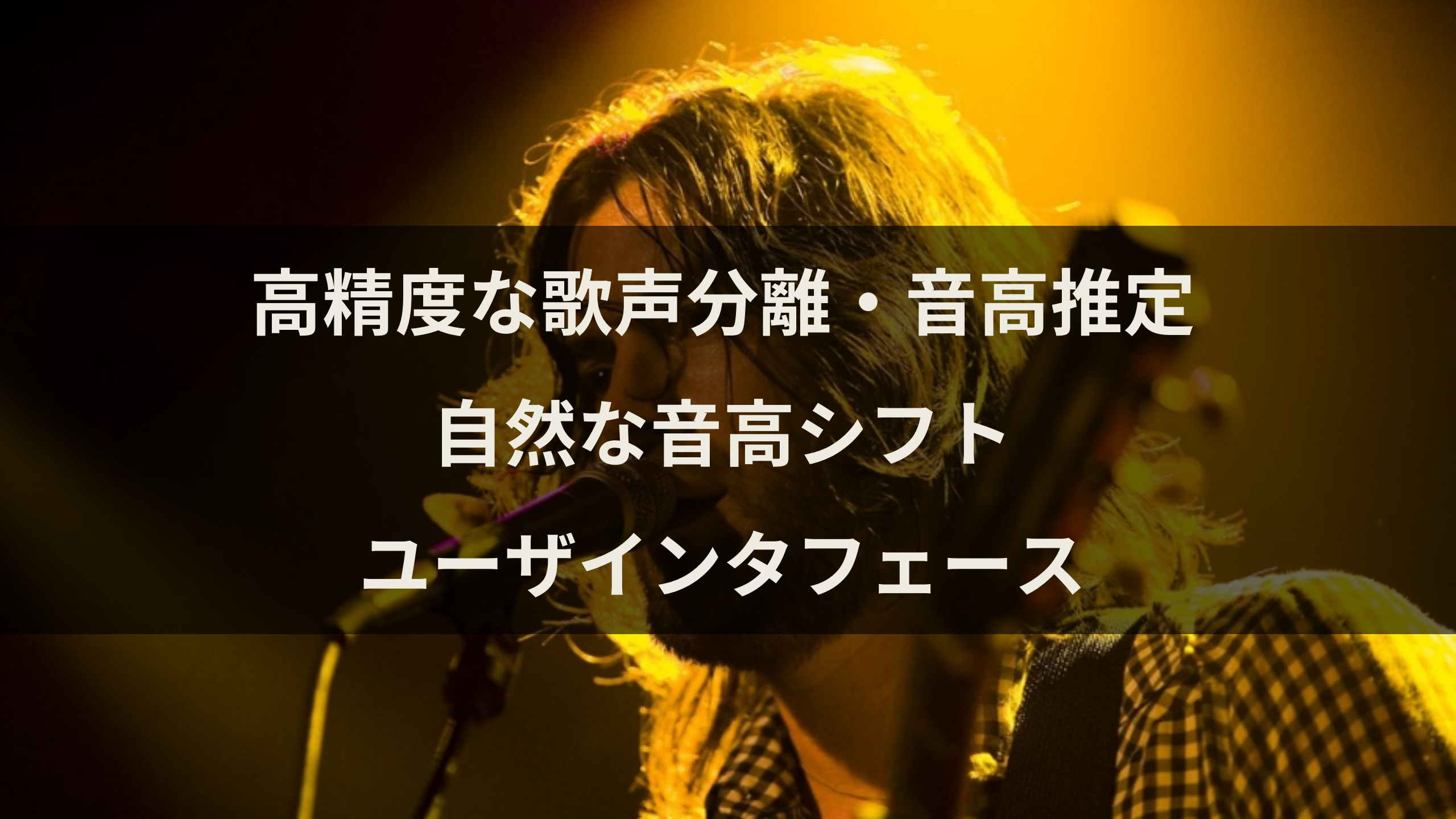
楽器パートごとに音量調節

歌声の音色・音高を編集

## 創作支援



	声質	音高
伴奏なし歌唱 (アカペラ)	統計的声質変換 [Toda 2014]	Phase Vocoder [Flanagan 1966]
	音声分析合成器 [Kawahara 2008, Morise 2011]	
伴奏つき歌唱 (市販CD)	W-PST法 [Fujihara 2010]	<b>本研究</b>

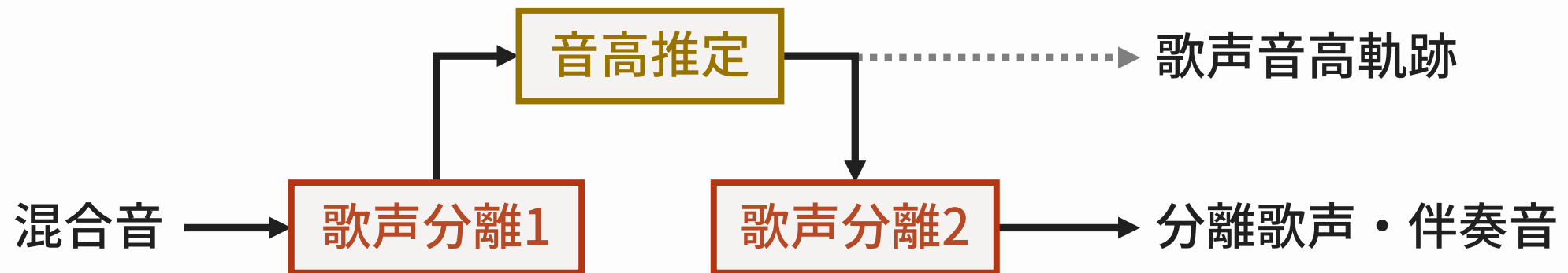


**高精度な歌声分離・音高推定**

**自然な音高シフト**

**ユーザインタフェース**



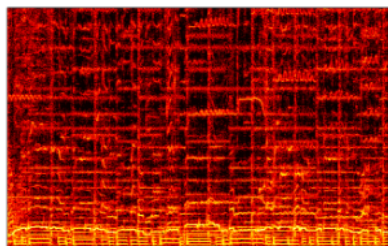


# RPCAによる 歌声分離

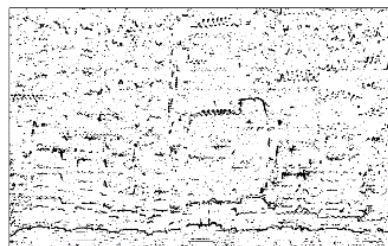
# 音高推定

# 音高を用いた 歌声分離

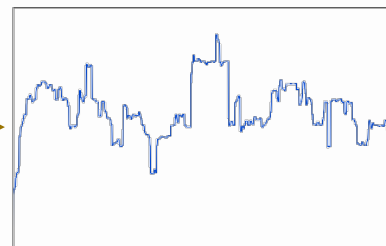
入力混合音



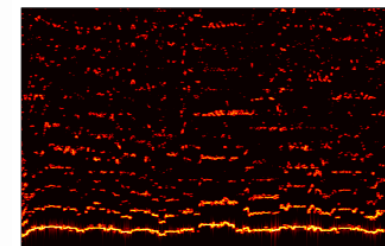
RPCAマスク



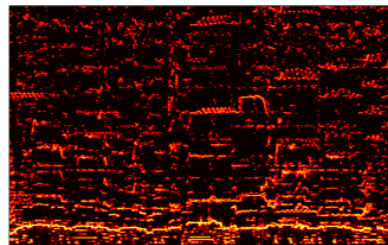
音高軌跡



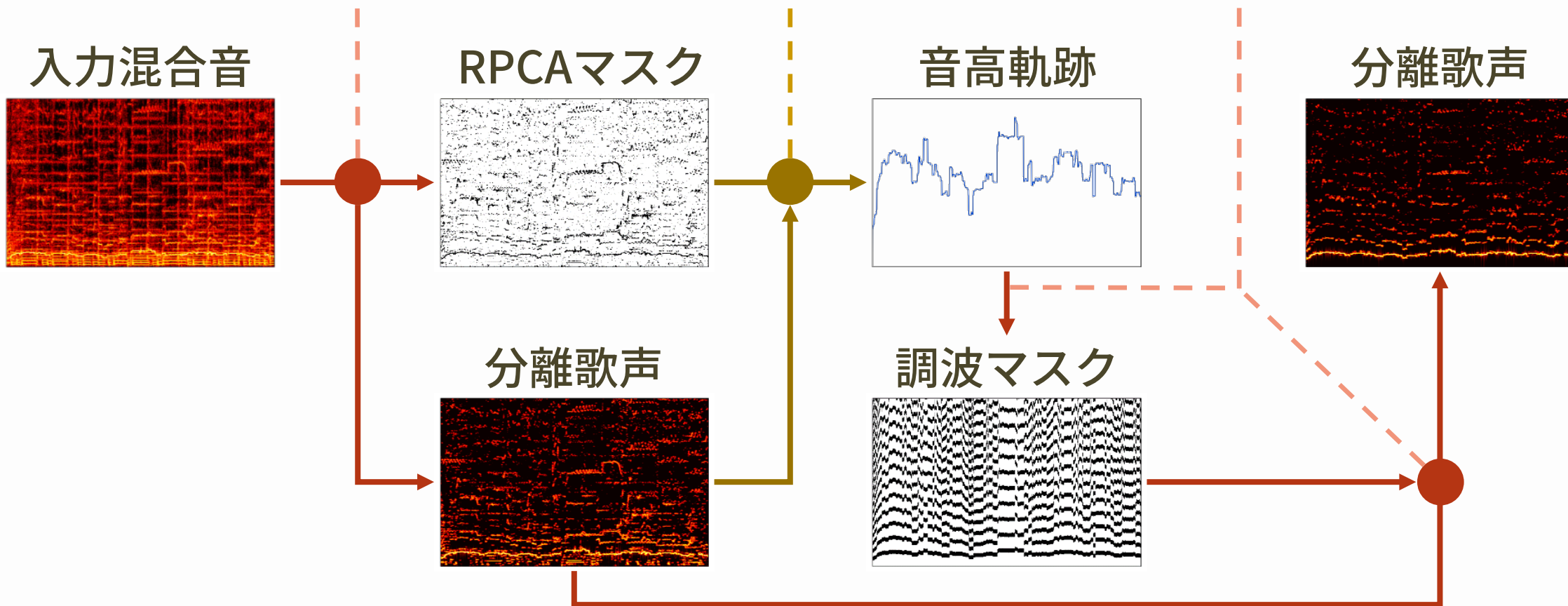
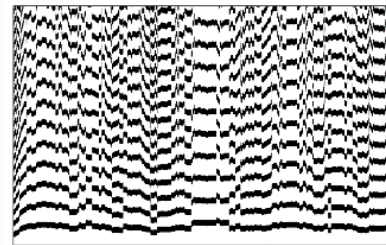
分離歌声



分離歌声

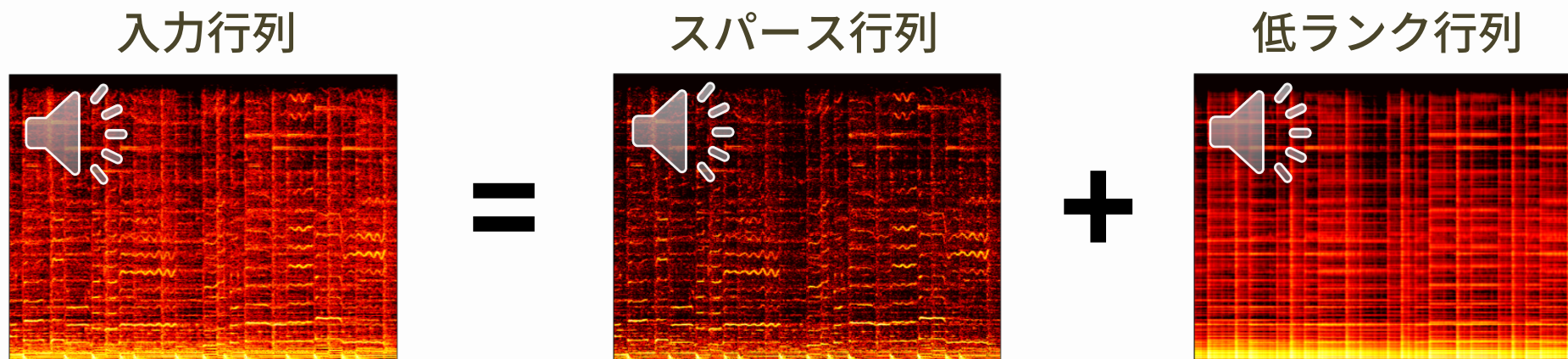


調波マスク



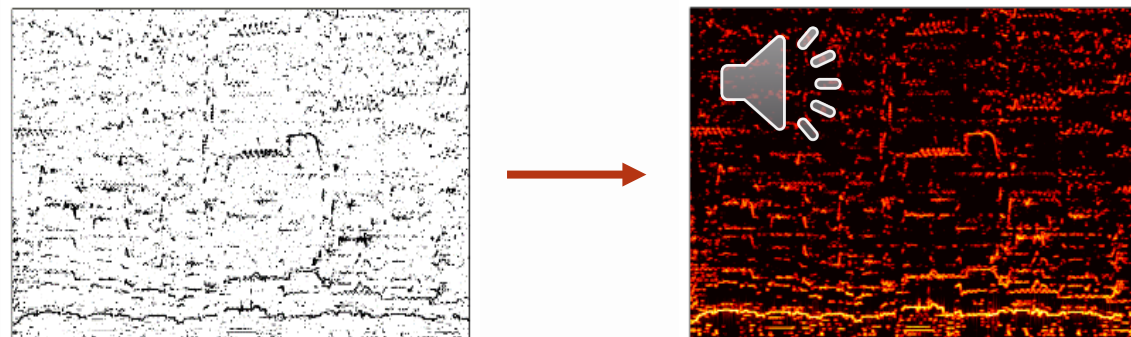


# ロバスト主成分分析による歌声分離 [Huang 2012]



スパース行列：稀にしか現れない成分 → **歌声**  
低ランク行列：繰り返し出現する成分 → **伴奏**

バイナリマスクの作成

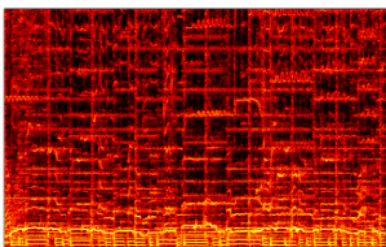


# RPCAによる 歌声分離

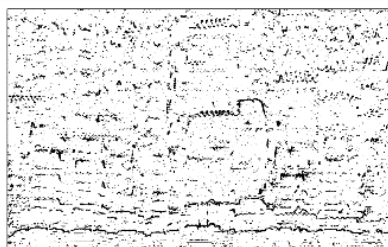
# 音高推定

# 音高を用いた 歌声分離

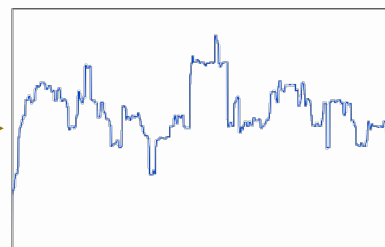
入力混合音



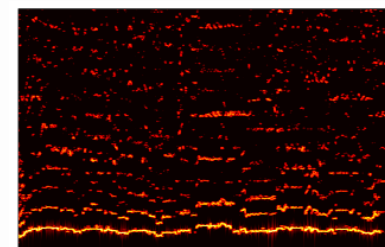
RPCAマスク



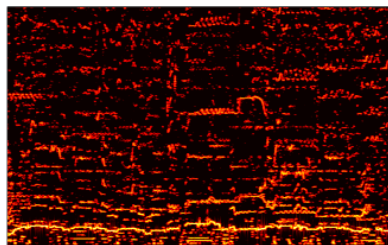
音高軌跡



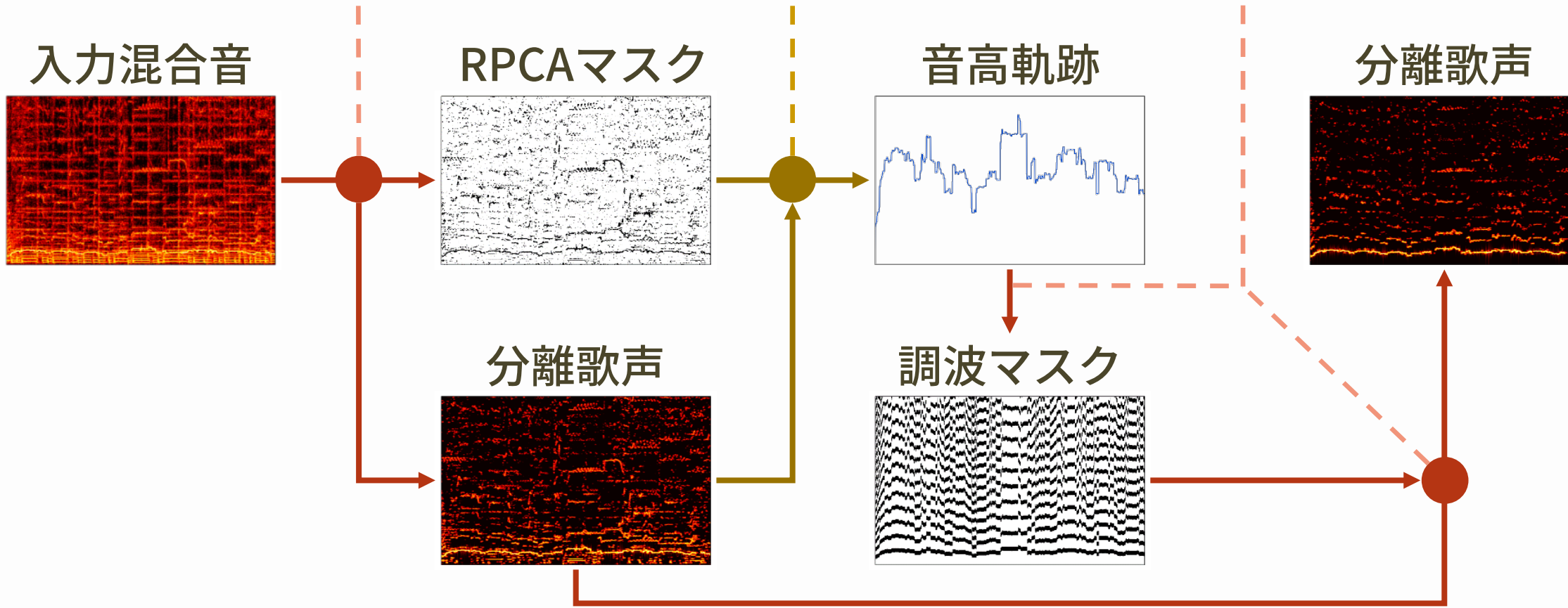
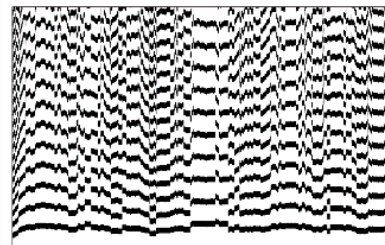
分離歌声



分離歌声

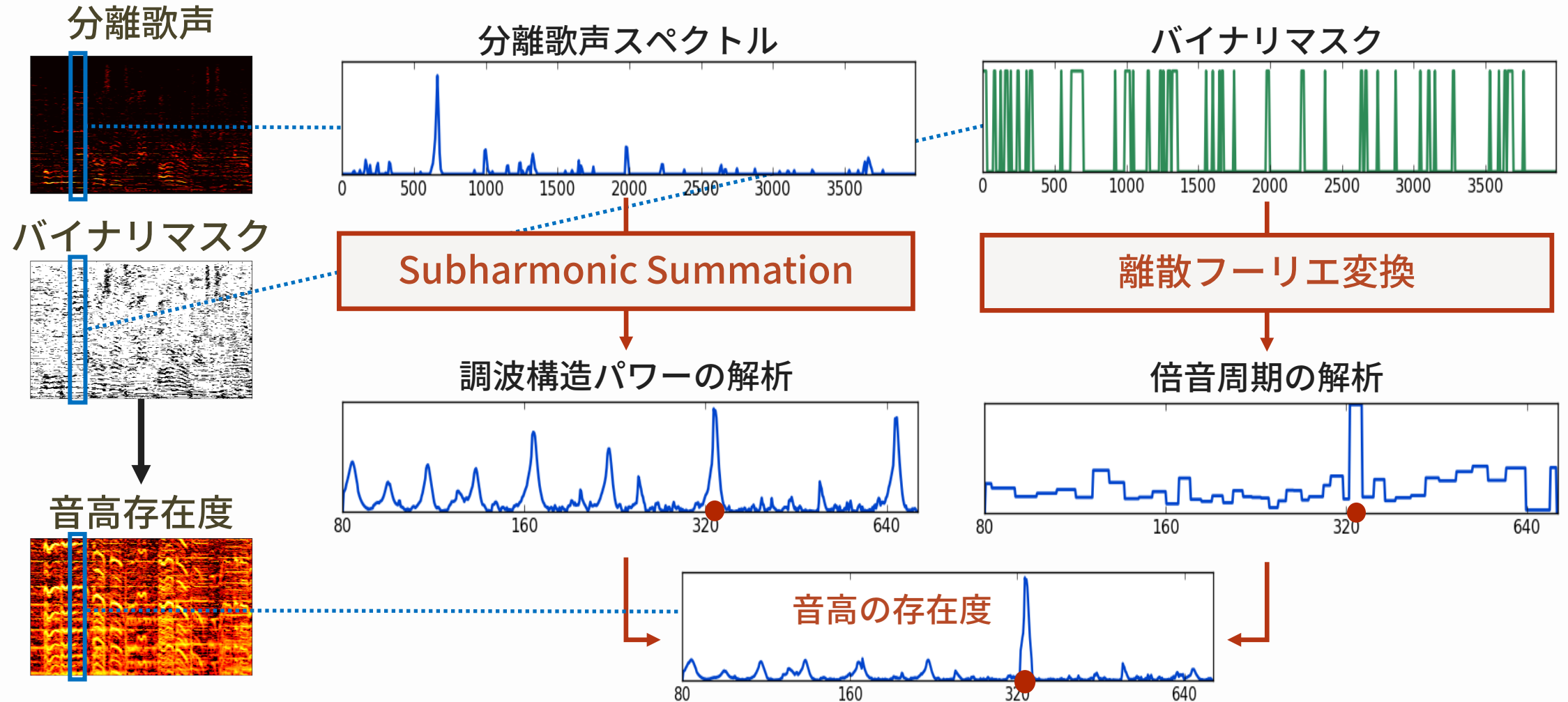


調波マスク



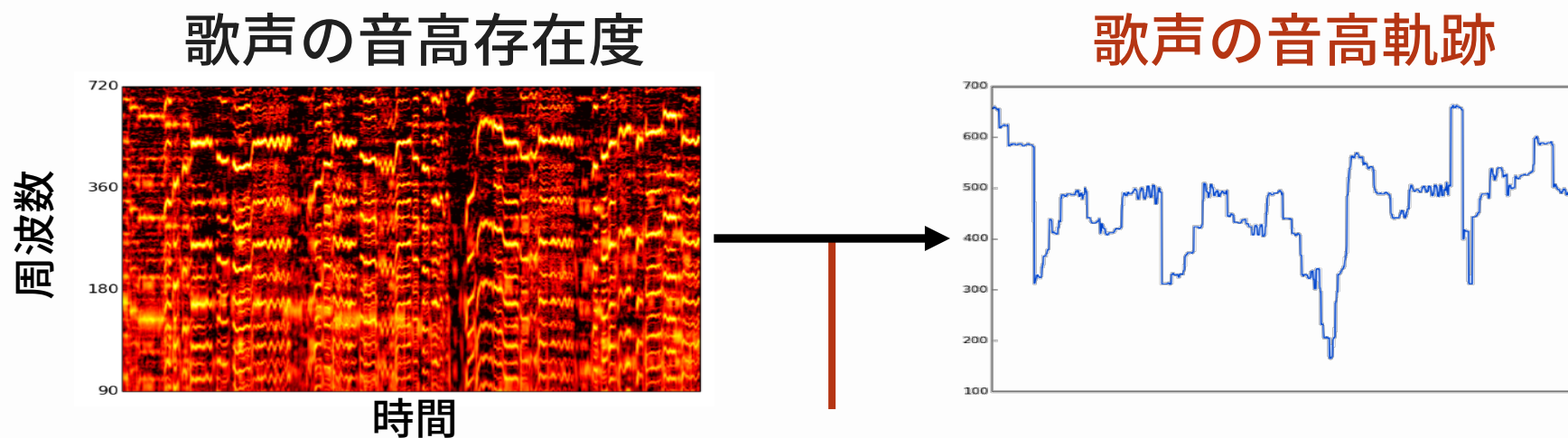


# 歌声の音高存在度の計算



# ビタビ探索による音高軌跡推定

動的計画法で最適経路探索問題を解く



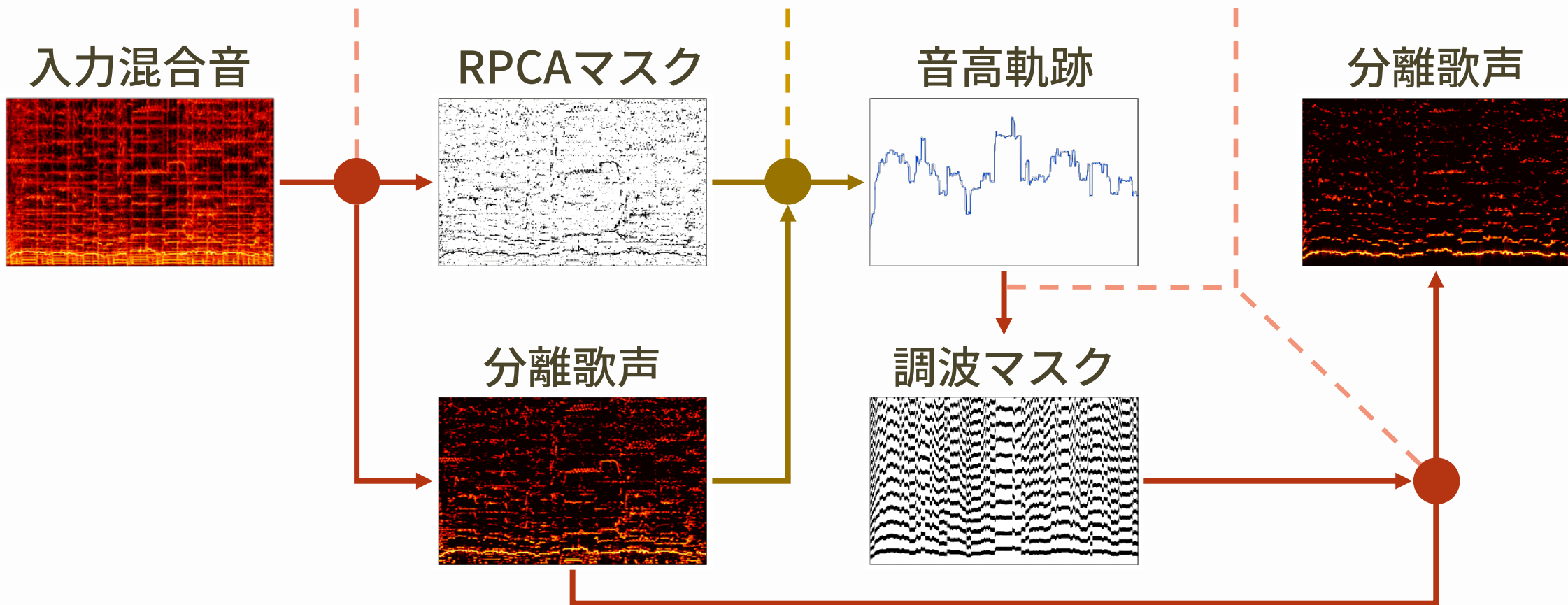
$$\hat{S} = \operatorname{argmax}_{s_1, \dots, s_T} \sum_{t=1}^{T-1} \{ \log a_t X_{\text{shs}}(t, s_t) + \log T(s_{t+1} | s_t) \}$$

音高が滑らかに遷移するという制約つき

# RPCAによる 歌声分離

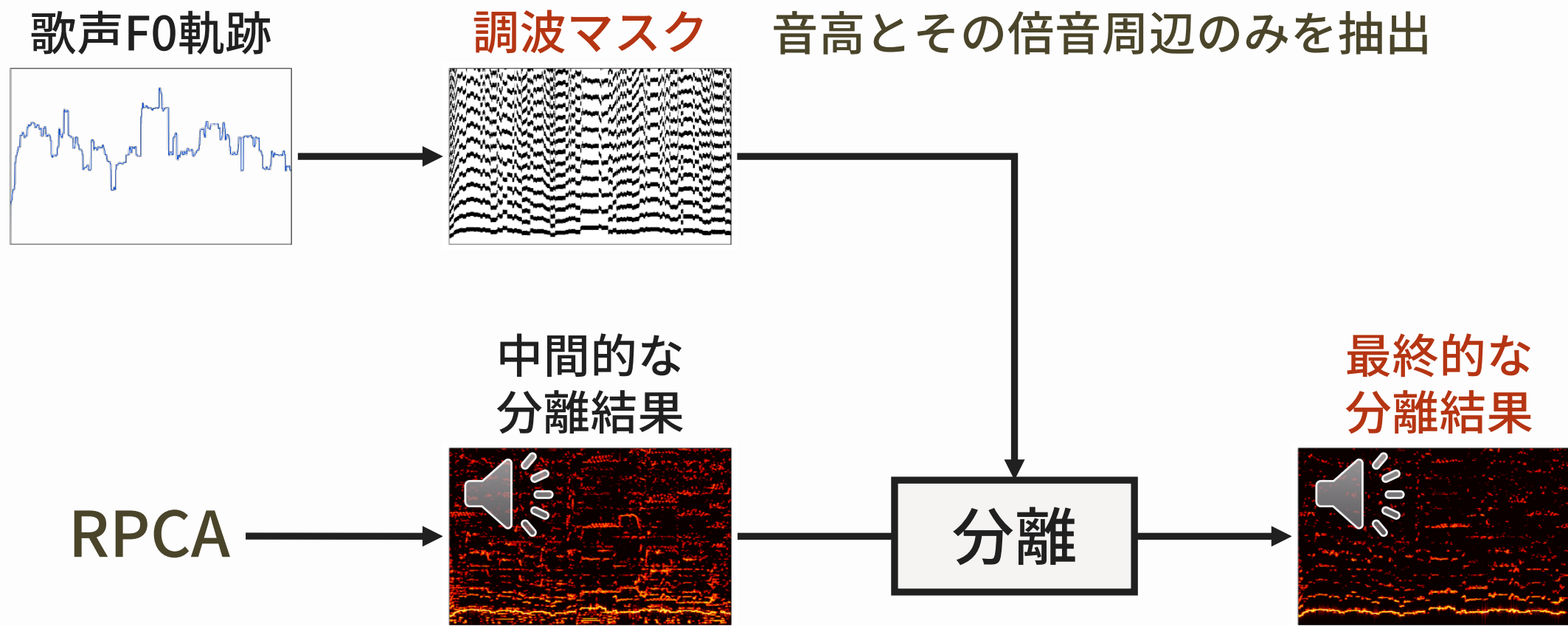
# 音高推定

# 音高を用いた 歌声分離



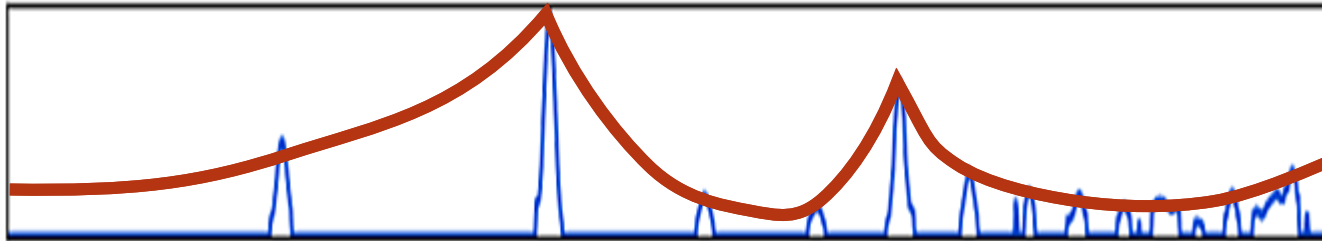


# 音高情報を用いて歌声分離精度を改善

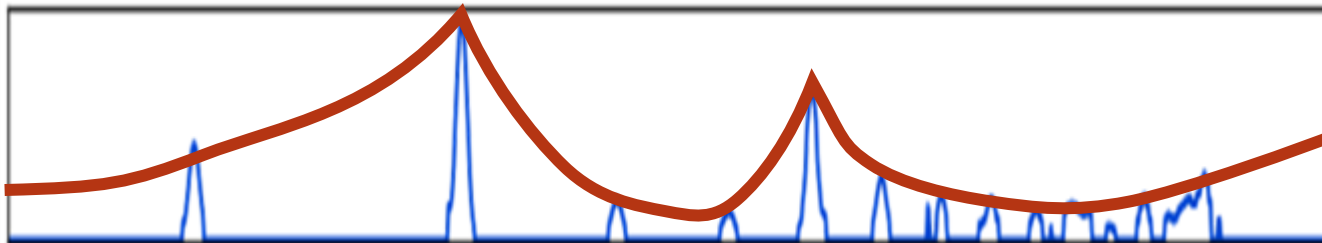


# 分離された歌声の音高操作

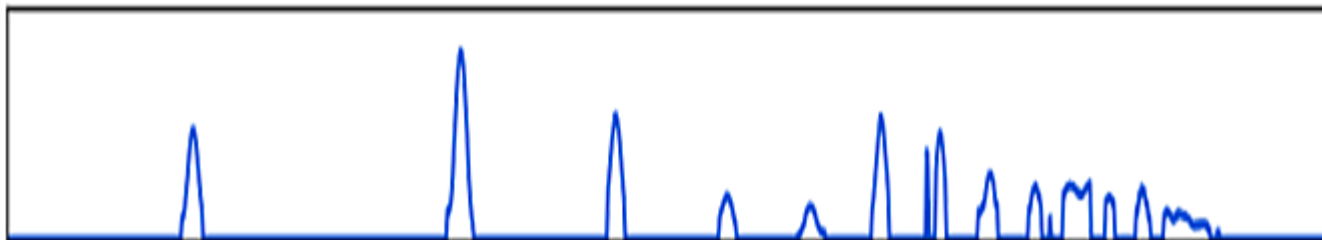
スペクトル包絡を保存して音色を補正



歌声スペクトル

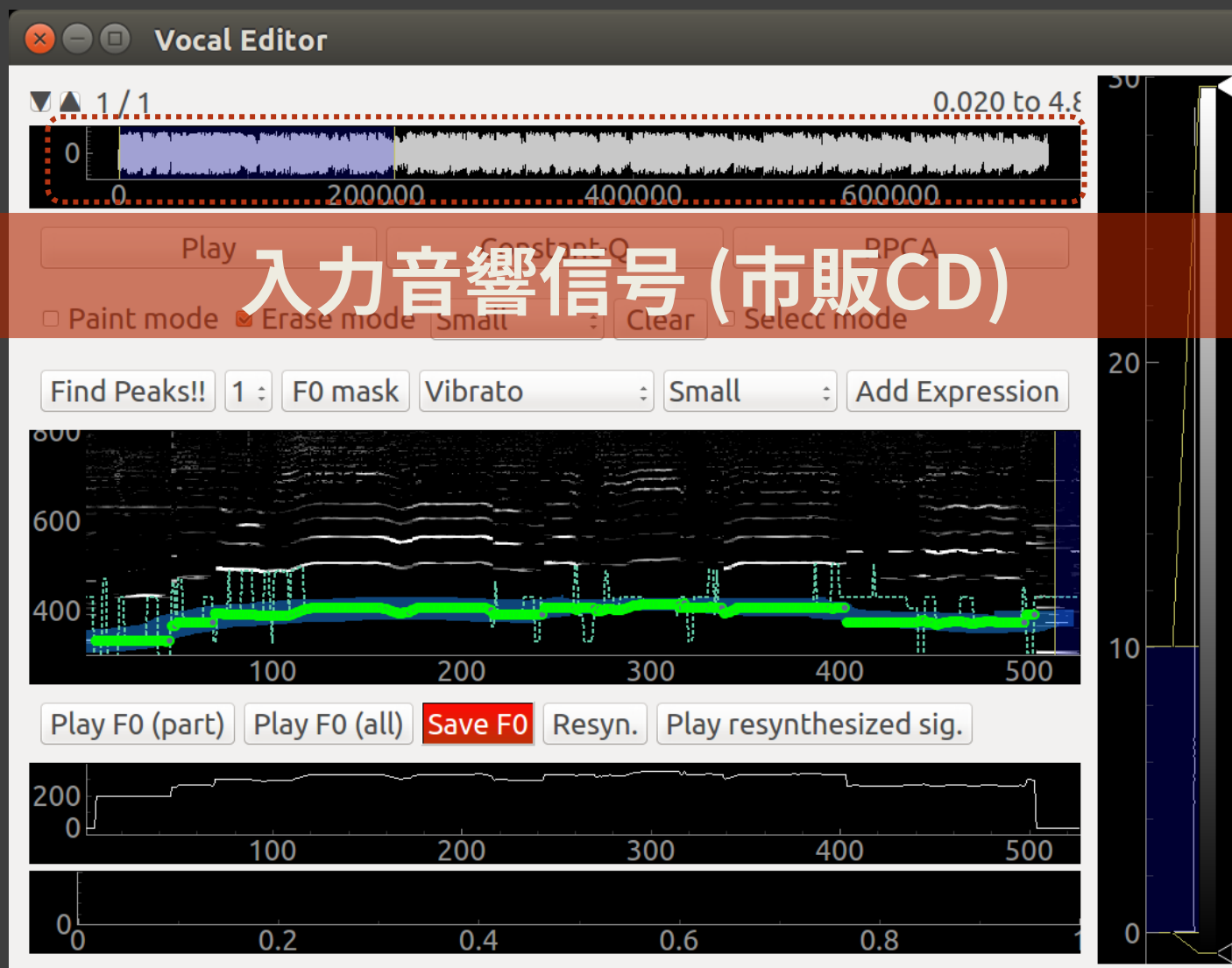


単純に音高シフト



音高シフト後に  
音色を補正

# 一般ユーザ向けインタフェースを開発



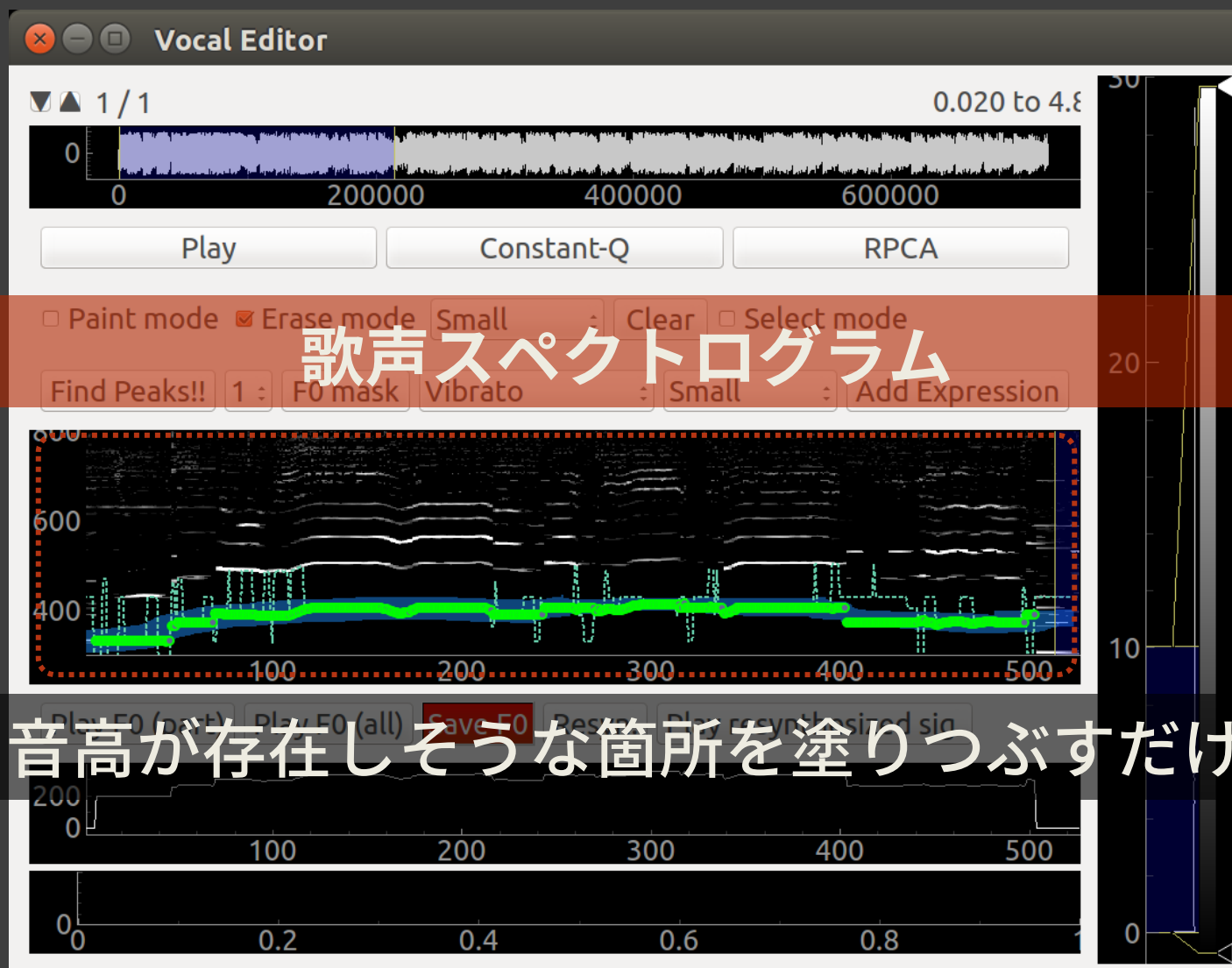


# ボタン一つで解析が可能

The screenshot displays the 'Vocal Editor' application window. At the top, there's a title bar with standard window controls and the text 'Vocal Editor'. Below it, a status bar shows '1/1' and a time range '0.020 to 4.8'. The main interface features a large waveform display at the top, with a red dashed box highlighting the 'Constant-Q' and 'RPCA' buttons. Below the waveform are several control buttons: 'Play', 'Constant-Q', and 'RPCA'. Further down, there are checkboxes for 'Paint mode', 'Erase mode', 'Small', 'Clear', and 'Select mode'. A 'Find Peaks!!' button is also visible. The central part of the interface shows a spectrogram with a green horizontal line indicating a frequency component. Below the spectrogram, there are buttons for 'Play F0 (part)', 'Play F0 (all)', 'Save F0', 'Resyn.', and 'Play resynthesized sig.'. At the bottom, there are two more plots: a line graph showing F0 over time and a smaller plot at the very bottom. A vertical volume or amplitude scale is visible on the right side of the interface.

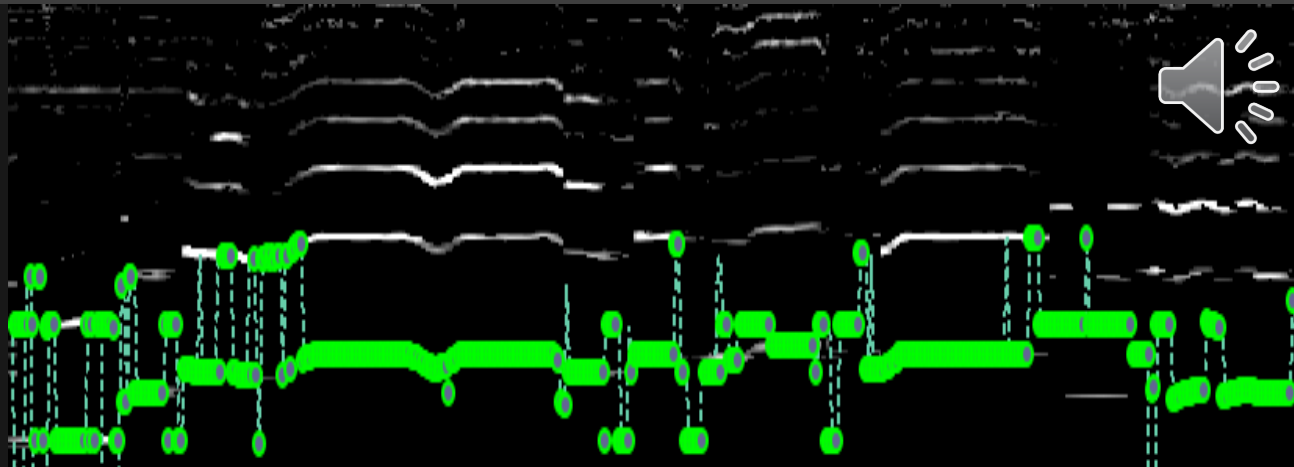
時間周波数解析  
RPCAによる歌声分離  
十分に高速に動作

# ユーザによる音高推定誤りの修正

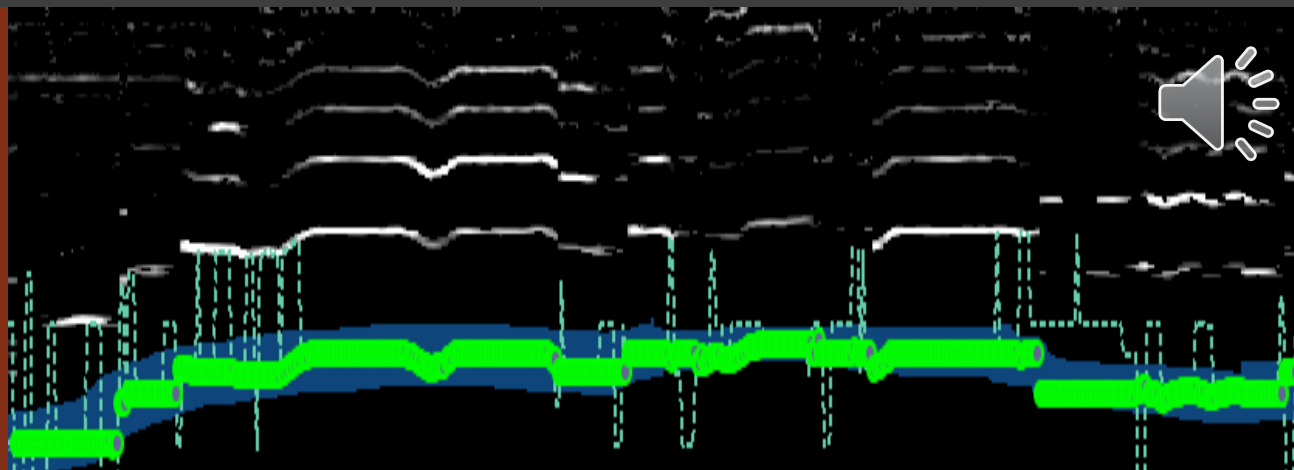


# 自動的に音高を再推定

自動推定された  
音高軌跡



ユーザ入力を用いて  
再推定された音高軌跡





# さまざまな歌唱表現テンプレート



歌唱表現の選択

ビブラート・こぶし  
グリスアップ・ダウン

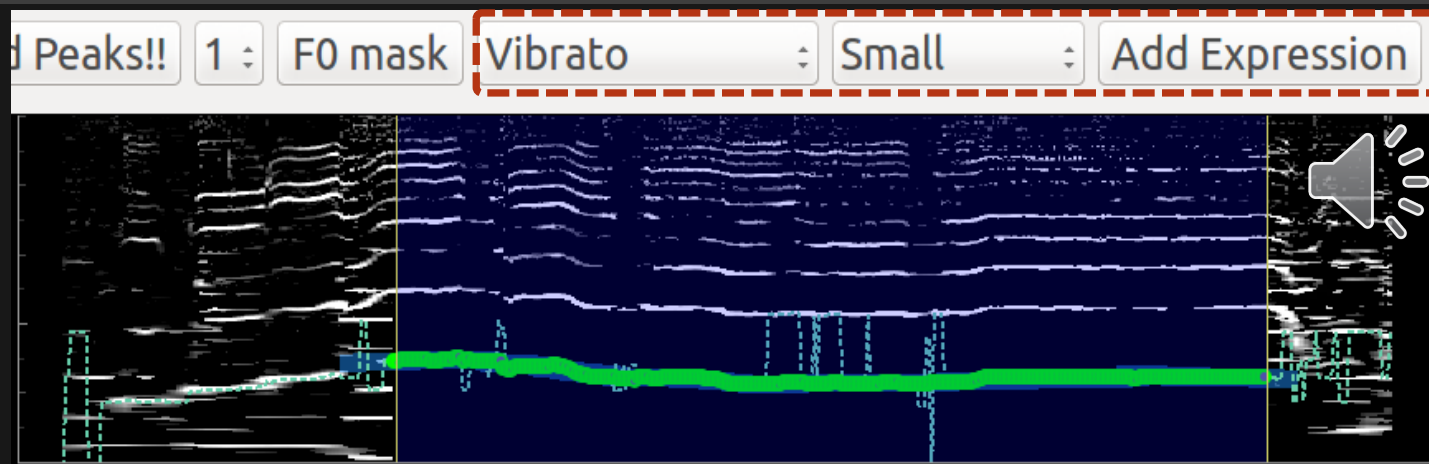
平滑化

移調

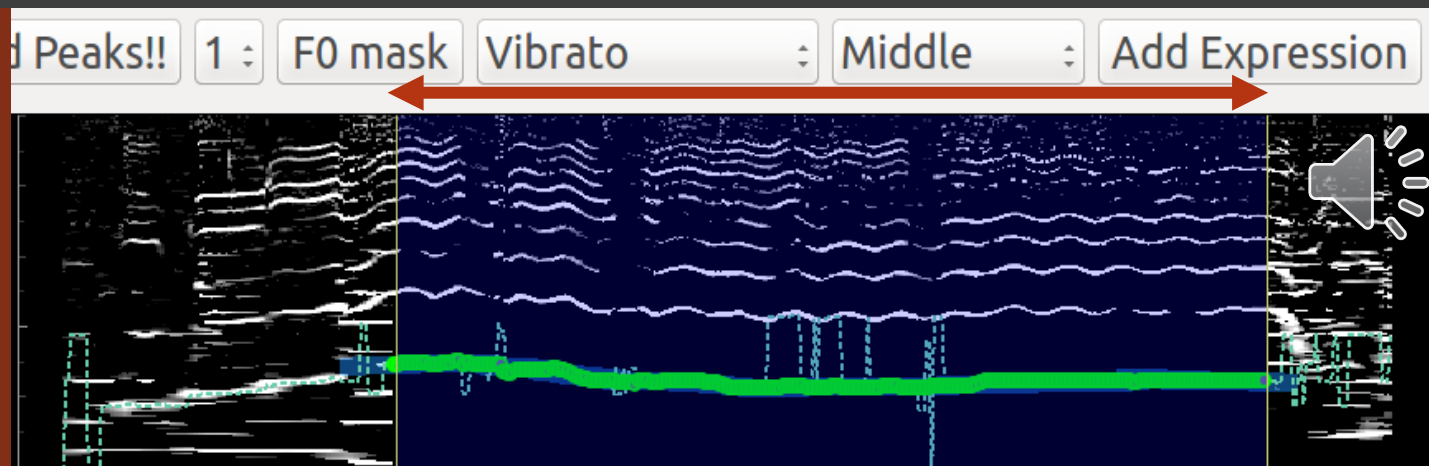
小  
中  
大  
特大

# 指定した範囲に歌唱表現を付加

オリジナル



ビブラート付加



# 評価実験

# 世界最高性能の歌声分離を達成

国際的な音楽認識コンテストMIREX 2014に参加

30秒の楽曲100曲 (データ非公開)

プログラムを提出して第三者の手によって評価

Deep Neural Network  
[Huang 2014]

Bayesian NMF  
[Yang 2014]

	HKHS	YC	LFR	RP	GW	RNA	JL	IIY
歌声	-1.40	-0.82	0.65	2.86	2.89	3.69	4.17	4.48
伴奏	0.35	-3.12	3.09	5.03	5.25	7.32	5.63	7.87

GNSDR [dB]



オリジナル

歌声

伴奏

そばかす /  
JUDY AND MARY



I'm in the mood for  
dancing / The Nolans



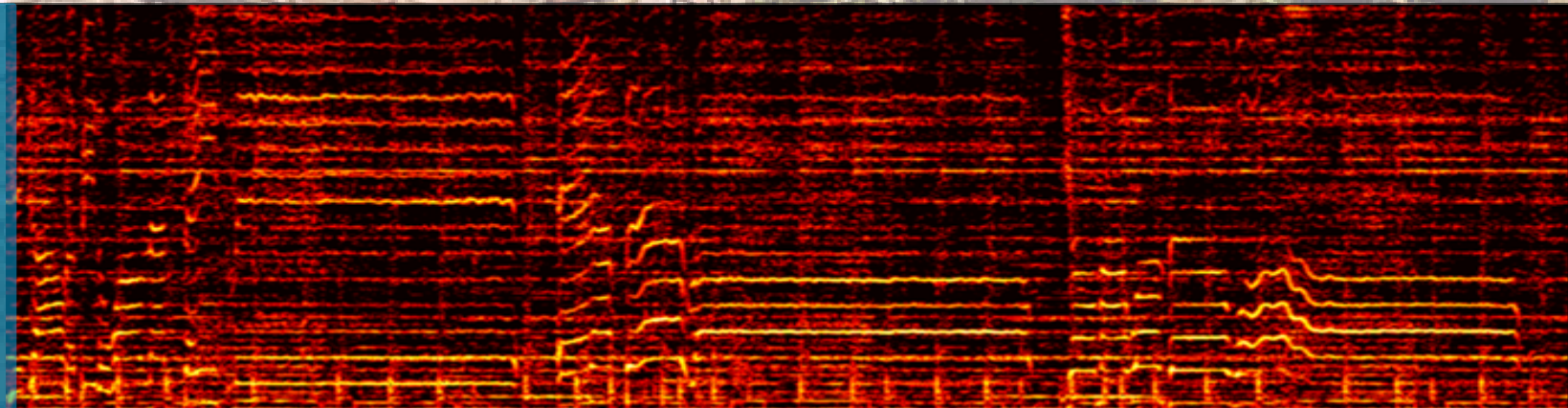
Morning Glory /  
Oasis



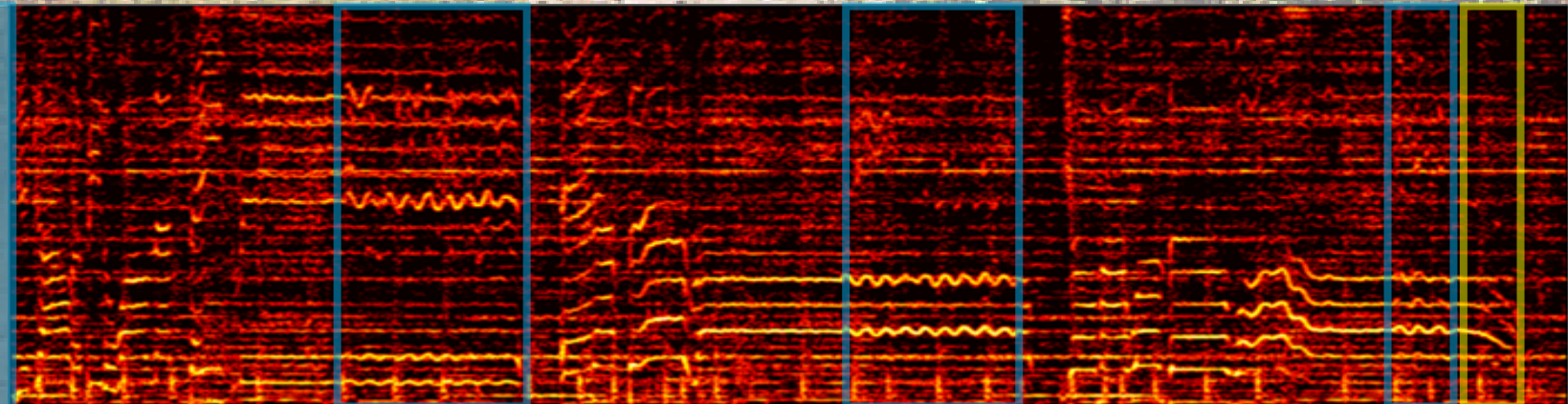
# ビブラートを付けて歌ってもらった

Wonderwall / Oasis

編集前



編集後

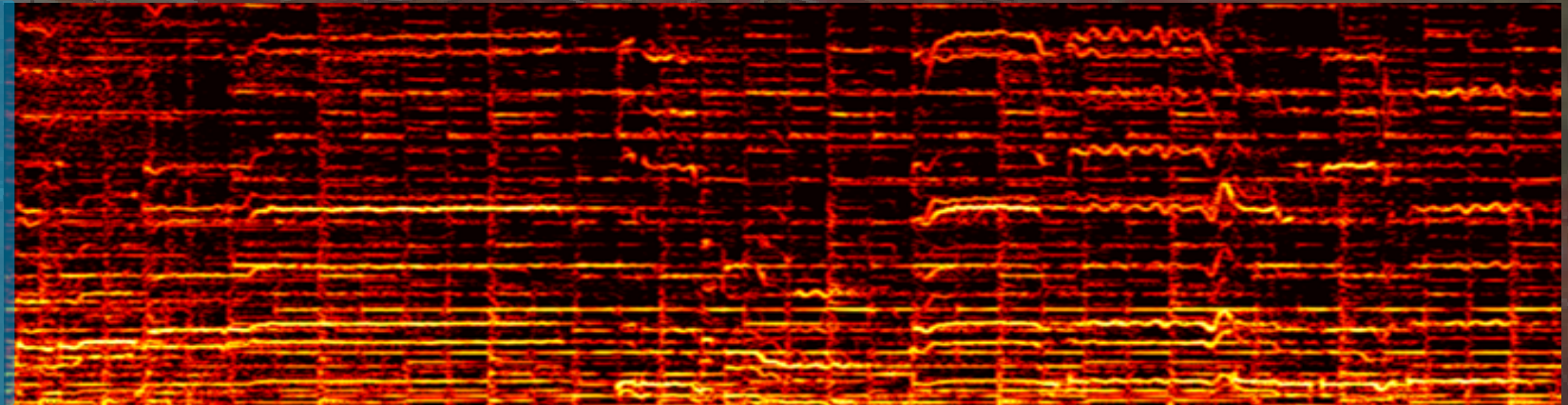




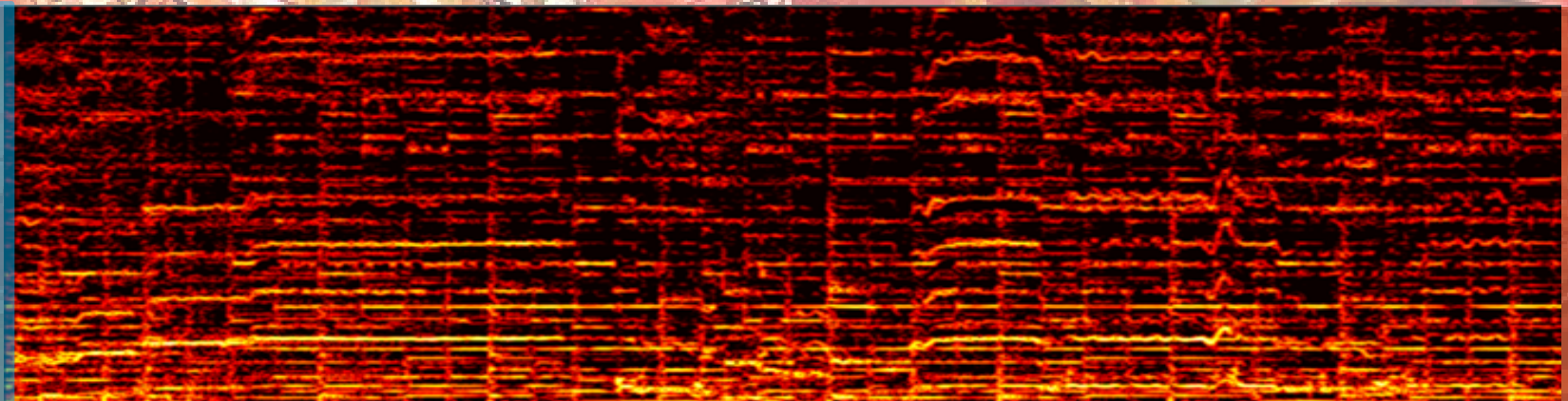
# ハモリを歌ってもらった (3音下げ&音量半分)

楓 / スピッツ

編集前



編集後

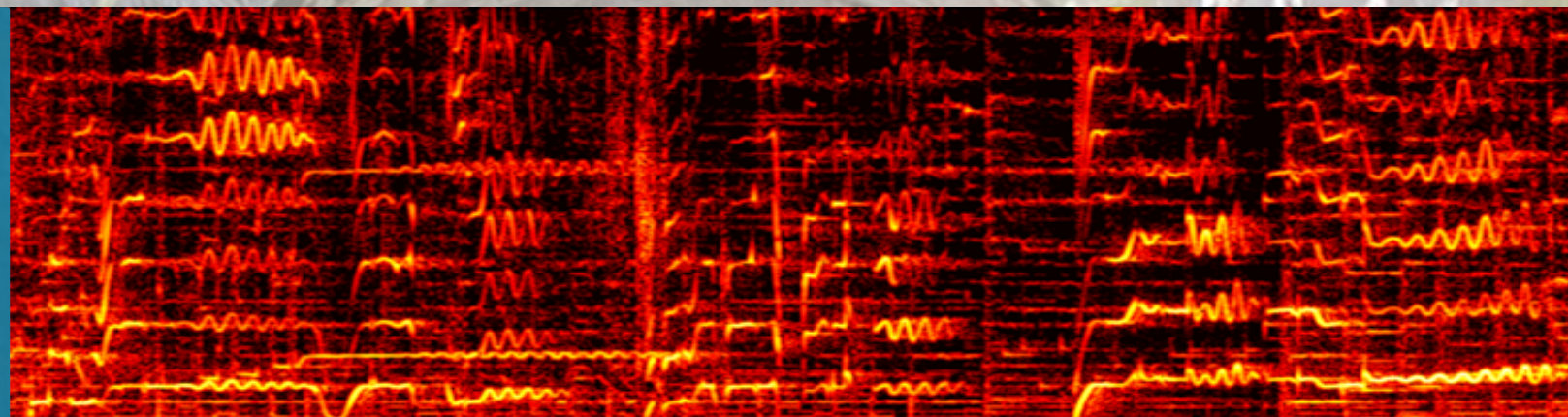




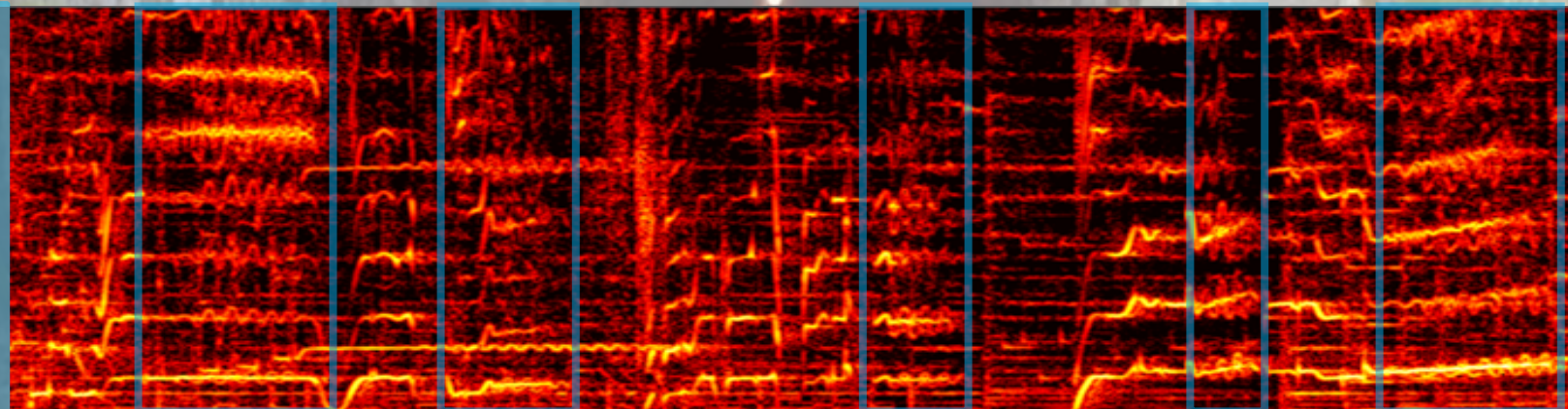
# 演歌をあっさり歌ってもらった (ビブラート除去)

天城越え / 石川さゆり

編集前



編集後





# 本研究の貢献

## 歌声分離・音高推定

歌声分離と音高推定の相互依存性に着目

双方のタスクで優れた精度を達成

## 歌声編集

分離された歌声のスペクトルに対して音高を操作

スペクトル包絡の保存による音色を補正

一般ユーザ向けGUIを実装

# 究極の目標に向けて

楽曲どうして歌手を入れ替えたい！

ある歌手が**自身の声質・クセ**でどの曲でも歌ってくれる

	声質	音高	音量
伴奏なし歌唱 (アカペラ)	統計的声質変換 [Toda 2014]	Phase Vocoder [Flanagan 1966]	音量軌跡モデル [小泉 2014]
	音声分析合成器 [Kawahara 2008, Morise 2011]		
伴奏つき歌唱 (市販CD)	W-PST法 [Fujihara 2010]	<b>本研究</b>	<b>今後の課題</b>

# 本研究の貢献

## 歌声分離・音高推定

歌声分離と音高推定の相互依存性に着目  
双方のタスクで優れた精度を達成

ポスター

C33

会場最奥

## 歌声編集

分離された歌声のスペクトルに対して音高を操作  
スペクトル包絡の保存による音色を補正  
一般ユーザ向けGUIを実装

技術提供・共同研究のご相談歓迎いたします



# 研究用ポピュラー音楽の分離

オリジナル

歌声

伴奏

RWC-P-2001 No.018



RWC-P-2001 No.033



RWC-P-2001 No.052



RWC-P-2001 No.066



RWC-P-2001 No.097



# グループ合唱の分離

オリジナル

歌声

伴奏

行くぜ！怪盗少女 /  
ももいろクローバー



ヘビーローテーション /  
AKB48



Nellie The Elephant /  
Toy Dolls

