

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

歌唱表現転写システム

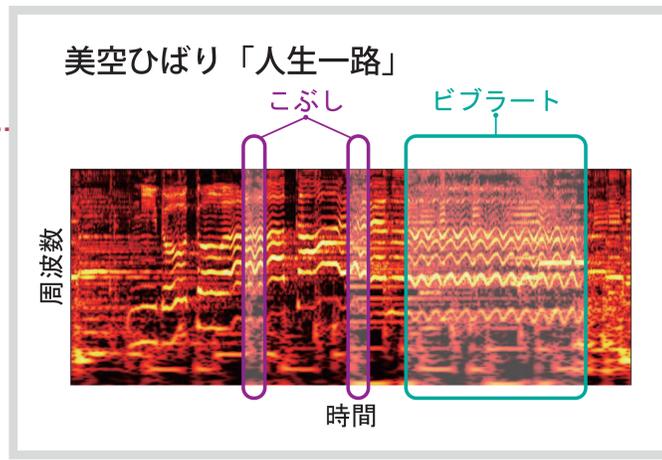
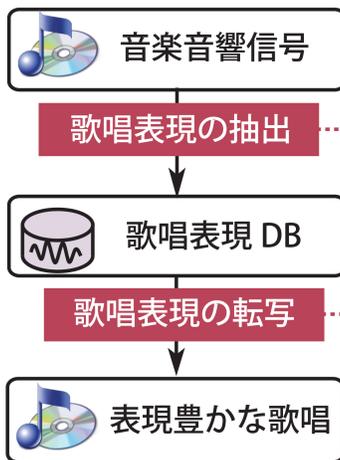
市販楽曲などの音楽音響信号に含まれる
歌声の歌唱表現を違う形にしたり無くし
たりして任意に**編集**するシステム

歌唱表現

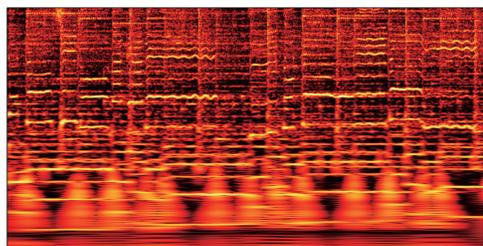
ビブラート、こぶし、グリッサンドなどの
歌声音高軌跡の特徴的な変動

こんなことができるように

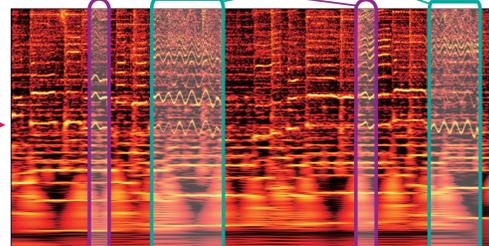
- ・「スピッツのグリスダウンが好きでたまらない。
全ての歌手はスピッツのように歌うべきだ！」
- ・「この曲のビブラートをもう少し小さくしたい！」



転写前



転写後



歌唱表現の抽出

歌声音高推定

- できるだけ簡単に
歌声の音高軌跡を知りたい -

時間周波数領域での
最適時系列探索問題
として定式化

- ・ 歌声 F0 軌跡の滑らかさを考慮 - マルコフモデル
- ・ 周波数探索範囲の制限

音楽音響信号 + 音高列

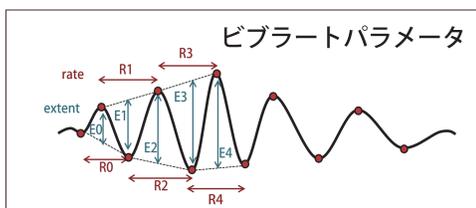


歌唱表現同定

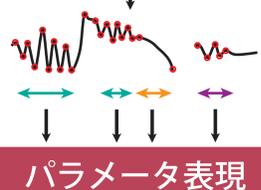
- 歌声音高軌跡から
歌唱表現を見つける -

パターンマッチング
により歌唱表現を同定

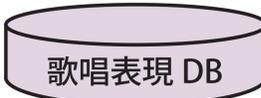
- ・ 各歌唱表現を操作可能な
パラメータ表現に落とし込む



特徴点検出 歌唱表現同定



パラメータ表現

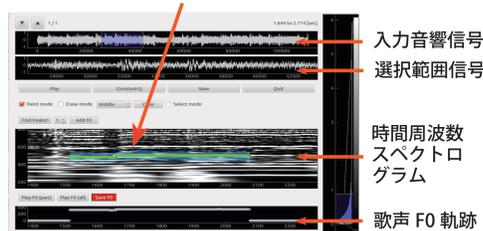


歌唱表現の編集

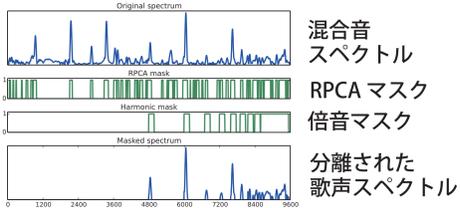
ユーザ主体の歌唱表現転写

- 任意歌唱の任意箇所へ
歌唱表現を転写できるシステム -

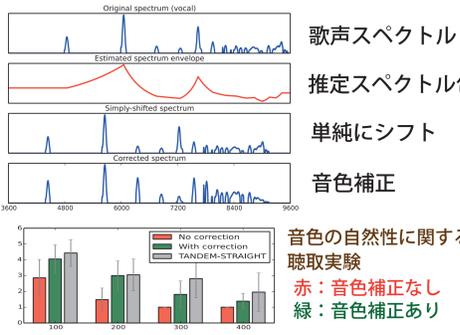
ユーザが F0 存在範囲・歌唱表現転写箇所を
スペクトログラム上で提示 (GUI 画面)



RPCA によるマスク & F0 を用いた倍音マスク

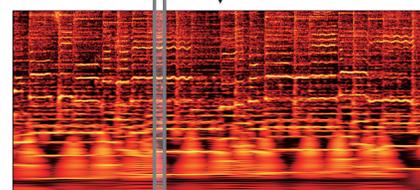


スペクトル包絡を用いた音色補正



音楽音響信号

定 Q 変換

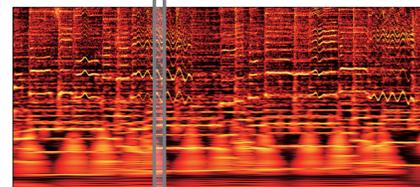


歌声分離

歌声スペクトル

音高シフト & 音色補正

伴奏
スペクトル



位相復元 & 定 Q 逆変換

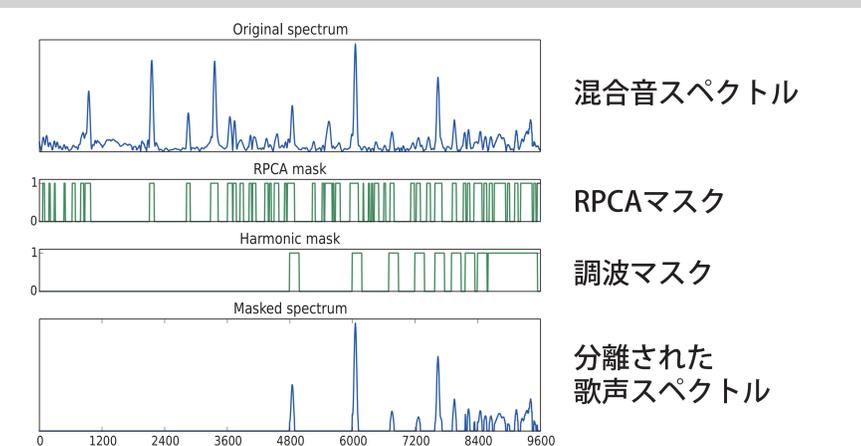
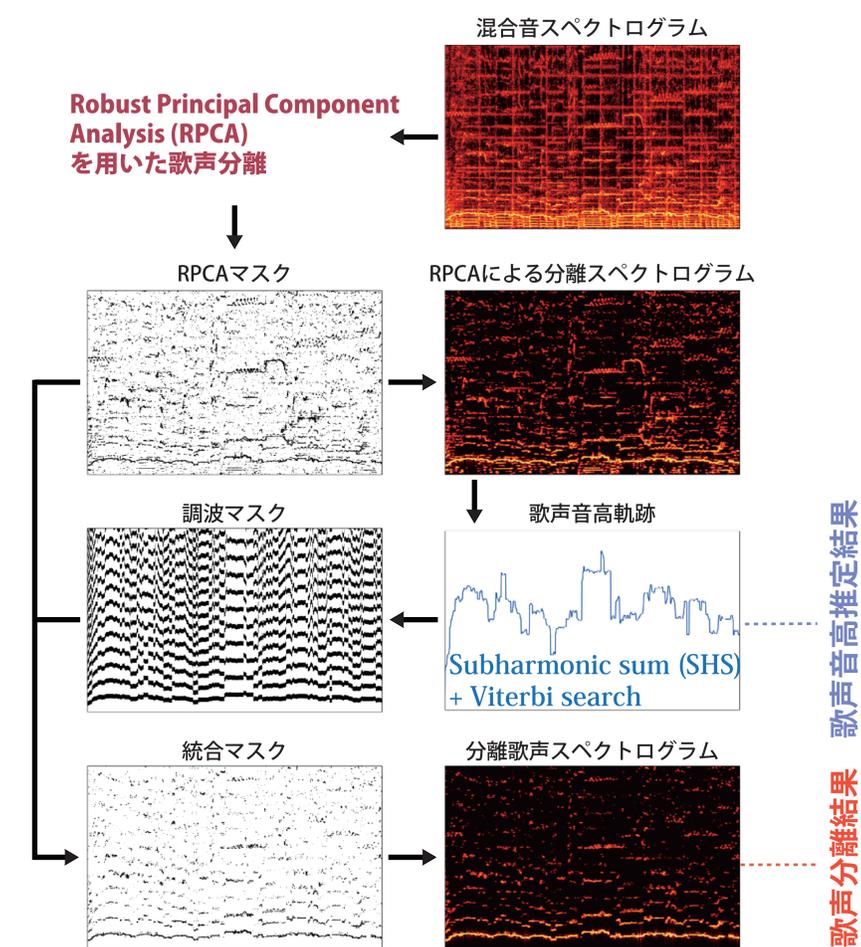
表現が転写された
音楽音響信号

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

～ 補足：歌声分離・歌声音高推定 ～

提案フレームワーク

歌声分離と歌声音高推定を相補的に行うフレームワーク

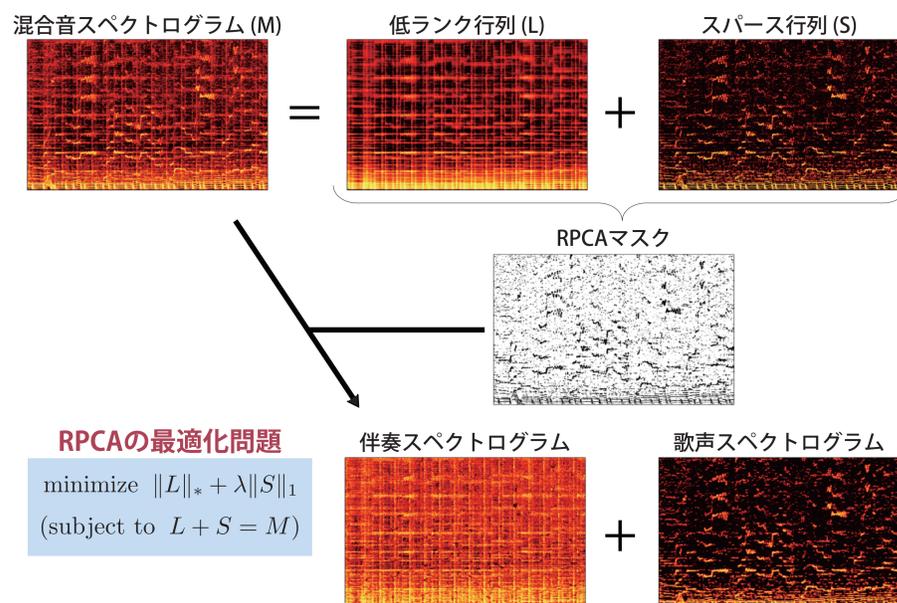


MIREX2014: 歌声音高推定

Submission	Overall accuracy	Raw pitch accuracy	Raw chroma accuracy	Voicing recall rate	Voicing false-alarm Rate	Algorithm
KD3	0.733	0.806	0.825	0.909	0.410	Spectral peak selection + Stream forming + Melody identification
KD1	0.722	0.793	0.813	0.864	0.325	Spectral peak selection + Tone modeling + Stream construction + Melody identification
DD1	0.714	0.722	0.750	0.860	0.296	Non-disclosure (doreso's proprietary algorithm?)
YJ2	0.540	0.748	0.786	1.000	1.000	Non-disclosure (Harmonic partial extraction + HMM?)
CWJ3	0.675	0.734	0.751	0.738	0.197	Pitch candidate extraction + Timbre- and loudness-based selection
IY11	0.634	0.728	0.772	0.851	0.362	Robust PCA + Subharmonic summation + Viterbi search
SL1	0.571	0.524	0.553	0.732	0.226	Acoustic model + Pitch evolution model + Viterbi search
LPSL1	0.406	0.405	0.466	0.678	0.359	Non-disclosure

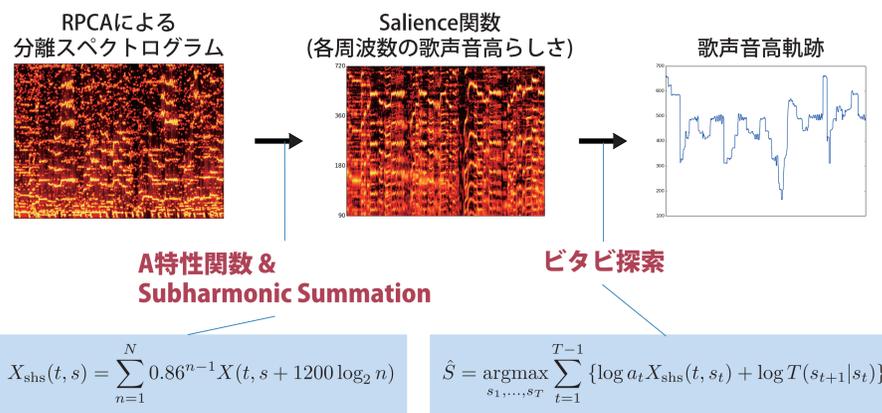
ロバスト主成分分析

ロバスト主成分分析 (RPCA) を用いた歌声分離



音高推定の定式化

RPCAにより分離された歌声から音高軌跡を推定



MIREX2014: 歌声分離

Submission	Voice GNSDR [dB]	Music GNSDR [dB]	Runtime [hours]	Algorithm
IY2	4.4764	7.8661	02	Robust PCA + F0 estimation (subharmonic summation)
IY1	4.2190	7.7893	02	Robust PCA + F0 estimation (subharmonic summation)
JL1	4.1564	5.6304	01	Harmonic/percussive/vocal separation (horizontal/vertical /sparse components)
RNA1	3.6915	7.3153	06	F0 estimation (harmonic sinusoidal modeling)
RP1	2.8602	5.0306	01	REPET-SIM (non-repeating component separation)
GW1	2.8861	5.2549	24	Bayesian NMF + Mel-frequency basis clustering
LFR1	0.6499	3.0867	03	Kernel additive modeling (non-repeating component separation)
YC1	-0.8202	-3.1150	13	Harmonic/percussive/vocal separation (spectral-temporal feature-based clustering)
HKHS1	-1.3988	0.3483	06	DNN/Deep RNN (trained on MIR-1K dataset)
HKHS2	-1.9413	0.5239	06	DNN/Deep RNN (trained on MIR-1K dataset)
HKHS3	-2.4807	0.1414	06	DNN/Deep RNN (trained on MIR-1K dataset)