Vol.2016-MUS-111 No.27
2016/5/21

# 歌声/調波楽器/打楽器音分離を応用したボーカル曲の性別変換再生

池宮 由楽[1,a)]

**概要**：楽曲に対してピッチシフトを行うことによりボーカルの性別を変換して再生する楽しみ方を提案する．これにより，男性楽曲は女性楽曲，女性楽曲は男性楽曲として楽しむことができる．しかし，楽曲を直接ピッチシフトすることには二つの問題がある．一つは歌声の音色が不自然になりうるという問題，二つ目が打楽器音がピッチシフトするのは不自然であるという問題である．これらを解決するために，歌声/調波楽器/打楽器音分離アルゴリズムを応用する．分離された歌声・調波楽器のみをピッチシフトし，また歌声はユーザが所望の声質となるようにフォルマントの調整を行うことで，自然な性別変換再生を楽しむことができる．本稿ではさらに，提案システムに適した分離アルゴリズムの提案を行う．

# Music Pitch Shifter based on Vocal / Harmonic / Percussive Source Separation

Yukara IKEMIYA[1,a)]

## 1. はじめに

音楽のデジタル化が一般的になるとともに，その再生方式は急速に多様化してきた．特に，WALKMAN や iPod に代表される携帯音楽プレイヤーには，ユーザが音楽の音色を調整して楽しむことのできるイコライザ・エフェクト機能が搭載され，さらに多機能なソフトウェア・アプリケーションも少なくない．このような音楽の加工は，主に音楽データに対するデジタル信号処理により実現されている．

CD 音源のように一度ミックスされた音楽を加工する場合，大別して 2 種類の加工方式がある．一つ目は，音楽信号全体を加工する方式である．フィルタリングによるイコライズ・リバーヴ付加や楽曲全体のピッチシフト・タイムストレッチがこれに当たる．二つ目は，音楽を部分的に加工する方式である．例えば，楽曲中で打楽器音のみ音色を変えたり [1]，各楽器パートの音量を変えて再生することができる [2]．これを行うための最も重要な技術的課題は音源分離である．音源分離は音響信号から特定の音を抽出す

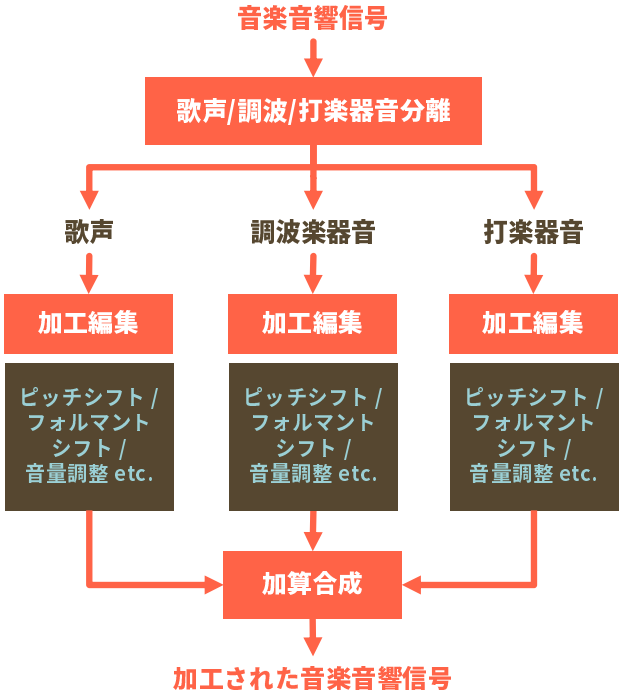


**図 1** 歌声/調波楽器/打楽器音分離に基づく楽曲加工システム

[1] 無所属，日本
[a)] yukara.ikemiya@gmail.com

る技術であり，音楽情報処理の分野では特に，楽曲からの歌声分離 [3], [4], [5], [6]，複数の楽器音の分離技術 [7], [8] が多く研究されている．

本稿では，音源分離を応用し，ボーカルの性別を変換して再生する視聴方式を提案する．性別変換再生の基本となるのは，楽曲のピッチシフトである．男性/女性ボーカルの大きな違いは歌唱音高の違いであるため，それぞれ適切にピッチシフトを行うことでボーカルの性別が変換されて聴こえるのだ．しかし，楽曲全体のピッチを変換して再生するのみでは二つの問題がある．

**1. 歌声の音色の不自然さ**

歌声をピッチシフトすると，音色に強く影響するフォルマント (スペクトル包絡) も同時にピッチシフトされる．このため，不自然にこもった声・甲高い声になることが多い．

**2. 打楽器音の音高の不自然さ**

打楽器はその性質上，音高の変化が相応しくない．特に，音高を下げる場合，必然的に一定帯域の高周波が消え去るため，楽曲全体がこもったような音となる．

前記の問題を解決するため，歌声/調波楽器/打楽器音を分離し，各パートを個別に加工することが可能なシステムを開発している (図 1)．分離された各パートに対して，ユーザはピッチシフト・音色調整・音量調整といった加工を施すことで，自分の好みの楽曲に仕上げることができる．特に，本システムの主目的である打楽器音以外のピッチシフトに適した分離アルゴリズムを提案する．

## 2. 歌声/調波楽器/打楽器音分離

多くの打楽器音分離手法 [9], [10] は，打楽器音スペクトルが時間方向に急峻であり，調波楽器音スペクトルが周波数方向に急峻であるという特徴を利用している．このアプローチは，各楽器音自体の特性を利用しているため楽曲構成によらず比較的頑健に分離できる一方，歌声のビブラートのようにピッチが急激に変化する調波構造を打楽器音として分離してしまうという問題がある．また，非負値行列分解を用いた手法 [11], [12] は，分解する基底やアクティベーションに各パートのスペクトルが満たすべき制約を課すことで，より精度良く分離しうるポテンシャルを示している．しかし，多くのパラメータをジャンルなどに応じた楽曲構成に合わせて調整する必要がある．

そもそも，歌声/調波楽器/打楽器音を一意に定義付けることができないことからも明白な通り，各パートを完全に分離することは不可能である．そのため，使用目的に対して適切な結果を得られる分離アルゴリズムを設計する必要がある．本稿で提案するシステムでは，打楽器音以外のパートのみピッチシフトすることを想定しているため，以下の項目を重視したアルゴリズムの設計を行った．

- 分離された打楽器音成分に他パートの成分ができるだけ混入しない．

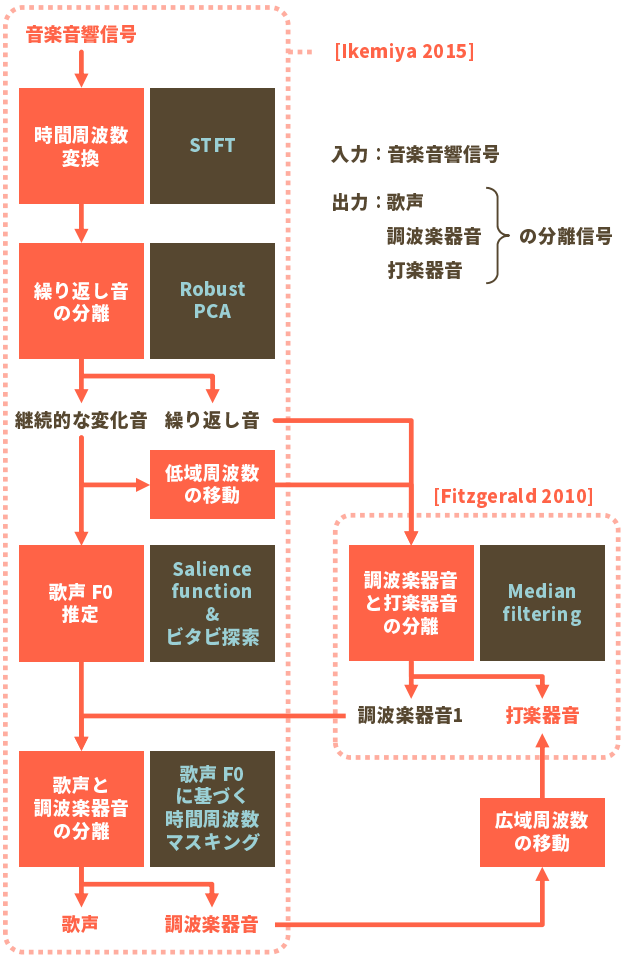


**図 2** 歌声/調波楽器/打楽器音分離

これは，打楽器音以外のパートがピッチシフトさせる成分とさせない成分に分離され，加工後の楽曲が大きく破綻することを防ぐためである．

提案する歌声/調波楽器/打楽器音分離 (VHPSS) アルゴリズムは，歌声分離 [5] と調波打楽器音分離 [9] を組み合わせた構成となっている (図 2)．Dobashi ら [2] はこの二つの手法を逐次的に行う VHPSS を開発したが，提案アルゴリズムはさらに分離打楽器音に他パート音が混入しづらい設計となっている．続く節で，アルゴリズム中の各パートについて説明する．

### 2.1 時間周波数変換

入力音楽音響信号はまず，短時間フーリエ変換 (STFT) により時間周波数領域のスペクトログラムへ変換され，振幅スペクトログラム上で VHPSS が行われる．

### 2.2 繰り返し音の分離

振幅スペクトログラムに対し Robust PCA を適用することで，楽曲の中で頻繁に繰り返している音 (繰り返し音) と継続的に変わり続けている音 (継続的な変化音) に分離

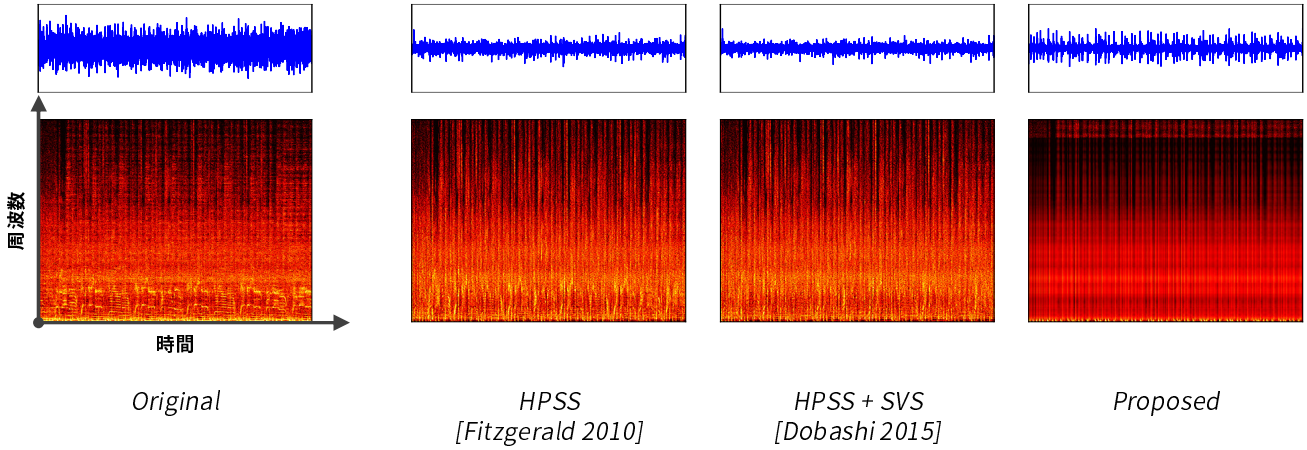


**図 3** 歌声/調波楽器/打楽器音分離の “*Smells Like Teen Spirit (Nirvana)*” に対する適用例．左からオリジナル楽曲，HPSS [9]，HPSS + 歌声分離 [2]，提案手法による分離打楽器音である．上下に信号波形と時間周波数スペクトログラムを示している．

することができる [3]．Robust PCA は入力行列を低ランクな行列とその差分であるスパースな行列に分離するアルゴリズムであり，時間周波数領域では繰り返し音と継続的な変化音が各行列の性質に適合している．

分離された繰り返し音には，例えば打楽器音やコードを弾くギター音などが含まれ，継続的な変化音には歌声やギターソロなどが含まれる．ただし，Robust PCA のスパース性の定義上，繰り返している音であっても，狭い周波数帯域にしかパワーを持たないベース音やバスドラム音はスパース行列として分離されてしまう．

### 2.3 歌声 F0 推定

分離された継続的な変化音 (2.2 節) から主旋律の歌声の基本周波数 (F0) を推定する．具体的には，時間周波数領域で歌声の F0 らしさを表す salience function を計算し，ビタビ探索により最適経路として F0 軌跡を推定する [5]．

### 2.4 調波/打楽器音分離

分離された繰り返し音 (2.2 節) から調波楽器音と打楽器音を分離する．このとき，Robust PCA でスパース行列へ分離された低域のドラム音を補正するため，低周波数成分 (例えば 200 Hz 以下) を繰り返し音のスペクトログラムへ移動する．調波/打楽器音分離はスペクトログラムの時間方向・周波数方向にメディアンフィルタを適用することによって行う [9]．歌声などは事前に継続的な変化音として分離されているため，打楽器音として誤り難い．

### 2.5 歌声/調波楽器音分離

Robust PCA の出力結果 (2.2 節) と推定歌声 F0 軌跡 (2.3 節) をから時間周波数領域でマスキング処理を行うことで主旋律の歌声を分離する [5]．このとき，分離された調波楽器音 (2.4 節) と継続的な変化音 (2.2 節) を比較することで適切なマスクを作成する．最後に，分離された調波楽器音の高周波数成分 (例えば 10 kHz 以上) を打楽器音スペクトログラムへ移動する．これは，楽器音の高周波数帯はスネアドラム・ハイハットなどの打楽器音のパワーが大部分を占めるという仮定に基づく．

## 3. 予備実験及びに検証

本稿で提案した VHPSS の有効性を確かめるための予備実験，及びに性別変換再生システムの検証を行った．なお，性別変換再生システムの完全な実装は行っていないため，現状可能な範囲での検証となる．

### 3.1 歌声/調波楽器/打楽器音分離の適用例

図 3 は市販楽曲 “*Smells Like Teen Spirit (Nirvana)*” に VHPSS を適用して分離した打楽器音の例である．メディアンフィルタを用いた打楽器音分離手法 (HPSS) [9], HPSS と歌声分離 [5] を組み合わせた手法 [2] による打楽器音分離結果との比較を示している．既存の二手法においては，歌声などが急激に変化している区間のスペクトルを打楽器として分離してしまい，時間波形を見ても明らかな通り打楽器音のアタック部が曖昧になり，ピッチを変更して再生する際に影響が出ることが考えられる．一方，提案法では分離打楽器音に対して歌声や調波楽器音の混入がほとんどみられず，特に打楽器音自体の分離割合も上がっている．

### 3.2 性別変換再生システムの検証

性別変換再生システムを簡易に実装し，現状の音質や課題の検証を行った．具体的には，分離された歌声・調波楽器音のピッチシフトを行い，歌声に対してはフォルマントシフトを行う．今回，ピッチシフトには Rubber Band

Library [*1] を利用した．また，分離した歌声に対するフォルマント (スペクトル包絡) の正確な推定は困難であるため，雑音による大きな推定誤りが起こりにくいケプストラムによる包絡推定を採用した．

分離パート毎の加工編集 (ピッチシフトなど) には，楽曲体験の拡張に対する大きなポテンシャルを感じた．楽曲全体をピッチシフトする場合では，各楽器の周波数帯域が変わることにより，その体感音圧比の変化や，耳障りな高周波の増加などの問題が起こりうる．提案システムにおいては，打楽器音を固定したり，各パートの音量を自由に変更でき，ユーザの好みに合わせて楽曲の微調整を行えることを確認した．

一方，音質面に関しては大きな課題が明らかとなった．特に，分離歌声に混入する調波楽器音の影響が大きく，フォルマントシフトなどを行った際に，一部の周波数のみが極端に持ち上げられ耳障りな音となる問題，フォルマント推定の誤差により二つの歌声が混じったような歌声になってしまう問題などが挙げられる．これらの問題に関しては，アルゴリズムの改善や 4 章で述べるインタフェース上の工夫などで解決を目指す予定である．

## 4. Future work

音楽の解析精度や編集後の音質面で実用性を考えると，全ての処理を全自動で行うことは現状難しい．現実的なアプローチとして，システムに対するユーザのインタラクティブな操作がある [13], [14], [15], [16]．例えば，解析する楽曲に関する簡単な情報を手動で入力したり，システムの出力結果の誤りを修正することで，より良い解析・合成結果を得ることができる．

本稿で提案したボーカル曲の性別変換再生システムにおいても，歌声/調波楽器/打楽器音分離部・楽曲の編集合成部の双方についてインタラクティブな操作が行えることが望ましい．各時間における歌声や打楽器音の有無を指定したり，楽曲のジャンルや楽器構成を入力すれば，それに適したアルゴリズムを選択し，解析性能を上げることができる．今後は，ユーザが入力・修正すべき要素の選別，操作方法の検討，インタラクティブな操作に適したアルゴリズムの検討を行っていく予定である．

## 5. おわりに

本稿では，楽曲のボーカル性別を変換して再生する楽しみ方を提案するとともに，そのシステムに適した歌声/調波楽器/打楽器音分離アルゴリズムを提案した．今後は，実用的なシステムの設計・実装に取り組む予定である．

**謝辞** 本研究は，京都大学における学生時代の研究内容を応用したものである．

[*1] http://breakfastquay.com/rubberband/

## 参考文献

[1] Nakamura, T., Kameoka, H., Yoshii, K. and Goto, M.: Timbre Replacement of Harmonic and Drum components of Music Audio Signals, *Proc. ICASSP*, pp. 7470–7474 (2014).
[2] Dobashia, A., Ikemiya, Y., Itoyama, K. and Yoshii, K.: A Music Performance Assistance System based on Vocal, Harmonic, and Percussive Source Separation and Content Visualization for Music Audio Signals, *Proc. SMC*, pp. 99–104 (2015).
[3] Huang, P. S., Chen, S. D., Smaragdis, P. and Hasegawa-Johnson, M.: Singing-Voice Separation from Monaural Recordings using Robust Principal Component Analysis, *Proc. ICASSP*, pp. 57–60 (2012).
[4] Rafii, Z. and Pardo, B.: REpeating Pattern Extraction Technique (REPET): A Simple Method for Music/Voice Separation, *IEEE TASLP*, Vol. 21, No. 1, pp. 71–82 (2013).
[5] Ikemiya, Y., Yoshii, K. and Itoyama, K.: Singing Voice Analysis and Editing based on Mutually Dependent F0 Estimation and Source Separation, *Proc. ICASSP*, pp. 574–578 (2015).
[6] M. McVicar, R. Santos-Rodriguez, T. D. B.: Learning to Separate Vocals from Polyphonic Mixtures via Ensemble Methods and Structured Output Prediction, *Proc. ICASSP* (2016).
[7] Itoyama, K., Goto, M., Komatani, K., Ogata, T. and Okuno, H. G.: Query-by-Example Music Information Retrieval by Score-Informed Source Separation and Remixing Technologies, *EURASIP JASP*, Vol. 2010, No. 1, pp. 1–14 (2011).
[8] Ewert, S., Pardo, B., Mueller, M. and Plumbley, M. D.: Score-Informed Source Separation for Musical Audio Recordings: An overview, *IEEE SPM*, Vol. 31, No. 3, pp. 116–124 (2014).
[9] Fitzgerald, D.: Harmonic/Percussive Separation Using Median. Filtering, *Proc. DAFx-10*.
[10] Tachibana, H., Ono, N., Kameoka, H. and Sagayama, S.: Harmonic/Percussive Sound Separation Based on Anisotropic Smoothness of Spectrograms, *IEEE/ACM TASLP*, Vol. 22, No. 12, pp. 2059–2073 (2014).
[11] Canadas-Quesada, F. J., Vera-Candeas, P., Ruiz-Reyes, N., Carabias-Orti, J. and Cabanas-Molero, P.: Percussive/Harmonic Sound Separation by Non-negative Matrix Factorization with Smoothness/Sparseness Constraints, *EURASIP JASMP*, Vol. 2014, No. 26 (2014).
[12] Park, J. and Le, K.: Harmonic-Percussive Source Separation using Harmonicity and Sparsity Constraint, *Proc. ICASSP* (2015).
[13] Goto, M., Yoshii, K., Fujihara, H., Mauch, M. and Nakano, T.: Songle: A Web Service for Active Music Listening Improved by User Contributions, *Proc. ISMIR*, pp. 311–316 (2011).
[14] Bryan, N. J. and Mysore, G. J.: Interactive Refinement of Supervised and Semi-Supervised Sound Source Separation Estimates, *Proc. ICASSP* (2013).
[15] Vincent, E., Bertin, N., Gribonval, R. and Bimbot, F.: From Blind to Guided Audio Source Separation: How models and side information can improve the separation of sound, *IEEE SPM*, Vol. 31, No. 3, pp. 107–115 (2014).
[16] Celemony: Melodyne とは？，Celemony（オンライン），入手先 〈http://www.celemony.com/ja/melodyne/what-is-melodyne〉（参照）．