Vol.2009-MUS-81 No.12
2009/7/30

スペクトルの時間変化に基づく音楽音響信号からの歌声成分の強調と抑圧

橘　秀幸[†1]　小野順貴[†1]　嵯峨山茂樹[†1]

本稿では，歌声と楽器音を両方含むような音楽音響信号から，歌声成分を強調，または抑圧する信号処理手法について述べる．歌声に相当する成分を検出するために，本稿ではスペクトルの時間変化に由来するスペクトログラムの特徴的な形状に着目する．歌声にはスペクトルの時間変化や旋律的な動きがあるため，スペクトルの形状が長時間一定であることはなく，またこれらの時間変化の影響で歌声のスペクトルは周波数軸方向にある程度の幅を有するという点で特徴的である．このような特徴をスペクトログラムの異方性という観点から捉えると，歌声と楽器音の滑らかさは異方的であり，異方的な信号を分離する手法を使って歌声と伴奏を分離することができる。本稿ではそのような手法を具体的に提案し，実際の音楽信号を用いた実験を行った結果，聴感上，歌声成分が強調/抑圧された信号が得られることを確認した．

Enhancement and Suppression of Vocal Components in Music Audio Signals Based on Temporal Variability of Spectra

Hideyuki Tachibana,[†1] Nobutaka Ono[†1]
and Shigeki Sagayama[†1]

We address a problem of enhancing or suppressing singing voice components in music audio signals. To achieve the purpose, we focus on peculiar spectral shapes of singing voice: they are not maintained unchanged for a while, and they occupy broad bandwidth, both of them is caused by spectral fluctuations and melodic nature of singing voice. When we regard those characteristic shapes as anisotropic smoothness of spectrogram, we can separate a music into singing voice and accompaniment, by applying a method which separates a signal into anisotropic components. In this paper, we propose a signal processing algorithm to enhance/suppress singing voice, based on those natures of spectral shapes of singing voice. We also conducted an auditory evaluation to confirm the effectivenes of the method using real music audio signals.

1. はじめに

我々が日常的に耳にする音楽の多くは，歌声によるメロディと，ギターやドラムなどの楽器による伴奏からなっている場合が多い．特に商業的な音楽では，歌声のパートに最も重点がおかれていることが多く，歌声のみを聴き取りたい，あるいは伴奏だけを取り出したカラオケが欲しいなどの需要は大きい．このため，歌声と伴奏を自動的に分離する技術には，大きな意義があると考えられる．

また，このような歌声強調/抑圧技術は，単にその出力を音楽鑑賞の対象として利用するのみならず，他の様々な音楽情報処理の前処理としても利用することができる．その典型例が，音楽情報検索[1] である．前述のような歌声に重点がおかれた楽曲の場合，歌声に関する様々な情報，たとえば歌手の声色や，歌声のメロディライン，歌詞の内容は検索の手がかりとして重要であると考えられるが，歌声と伴奏の混合信号からこれらの情報を直接抽出するのは難しく，先に歌声を強調しておくことが有効であると考えられる．

歌声強調/抑圧に類似する技術として，従来から，雑音下の音声（話し声）を対象とした，音声と背景雑音を分離する音声強調技術が研究されている．しかし，歌声強調と音声強調には，いくつかの異なる点がある．その一例として,「雑音」に想定されている性質の違いが挙げられる．音声強調においては，雑音には，白色性や，定常性，音声との無相関性などが仮定されることが多いのに対し，歌声強調おいて「雑音」に相当するのは，ギターやドラムなどによって演奏される伴奏であり，これは音声強調で仮定されるような雑音の性質は満たさない．すなわち，通常の音楽では伴奏が白色定常雑音でないのは明らかであるし，伴奏は歌声と同一の和音やリズムパターンに従って演奏されるため両者は無相関ではない．このように，歌声強調には従来の音声強調手法のみでは扱いが困難な問題も含まれている．

歌声強調などに関するこれまでの研究としては，2), 4) などがある．これらでは，歌声区間の検出と，その区間における基本周波数推定，さらにその情報を用いた歌声の分離という，3 段階の処理によって歌声を強調している．このような方法の場合，各要素技術の性能を向上させることで，全体の性能を向上させることができる反面，ある段階の処理がボトルネックとなって全体の性能が低下する懸念もある．たとえば，二重唱（デュエット）などで

†1 東京大学大学院情報理工学系研究科
Graduate School of Information Science and Technology, The University of Tokyo

はそれぞれのパートの基本周波数を推定するのが非常に難しいが，実際にはこのような楽曲も多い．このため，このような方針とは別の角度からのアプローチも検討する必要がある．

そこで，歌声のスペクトルの局所的な性質のみに着目して，直接的に信号を操作するフィルタリングのような手法が有効であると考えられる．歌声のスペクトルに局所的に現れやすい性質としては，ヴィブラートなどに由来するスペクトルの特徴的な形状を挙げることができる．これは歌声のスペクトログラムの滑らかさの方向は，ヴィブラートや旋律的な動きなどの影響によって，楽器音とは異なる方向になりやすいという性質である．スペクトログラムの滑らかさの異方性に基づいた楽器音分離手法としては，我々の研究室でこれまでに，ギターなどコード楽器音 + 歌声と，打楽器とを分離する，調波打楽器音分離[5)–8)] という手法を提案している．本稿では，この手法を少し条件を変えて用いることで，同様な原理でコード楽器音と歌声も分離することができることを示す．また，この手法を従来からの打楽器分離と組み合わせることによって，歌声を強調した信号や，抑圧した信号を得ることができる．

2. 歌声の持つ性質：スペクトルの時間変化

歌声には，基本周波数の不規則な微細変動や，ヴィブラートと呼ばれる 5–8Hz 程度での準周期的な変動がある[9)]．また，歌声は旋律的に歌唱される場合が多く，歌声にはしばしば旋律的な性質も付随する．これに対し，ギターなどのコードの音は，発音の瞬間に急峻な振幅の変化がある以外は，一度鳴った後は比較的長時間音が鳴り続け，特にいわゆる白玉コードのような場合にこの性質は顕著である．このことから，歌声は非定常的な音，コードを演奏する楽器の音は定常的な音とみなすことができる．なおこれ以降，後者のような音を「コード楽器音」と呼ぶ．

ところで，定常信号はどのような時間区間の短時間スペクトルを観測しても，ほぼ同一のスペクトル形状を示す．このため，コード楽器音を連続的に短時間フーリエ変換して得られたスペクトログラムを観察すると，比較的長時間一定のスペクトル形状を保たれるため，時間軸方向に滑らかな形状で表現される．これに対し，非定常信号の短時間スペクトル形状は，スペクトル解析を行う時間区間に依存する．したがって歌声のスペクトログラムでは，長い時間幅で見たときに一定のスペクトル形状は保たれず，時間軸方向に滑らかには表現されない．

また，歌声の基本周波数や振幅の変動は，周波数変調 (Frequency Modulation, FM) や振幅変調 (Amplitude Modulation, AM) とみなされることがある．これに対しコード楽器音は，発音の瞬間や，ギターやピアノなどでは振幅が少しずつ減衰する点や，いくつかの特別な奏法で演奏した場合などを例外として，ピッチや振幅は大きくは変動しないため，比較的変調の浅い，無変調波とみなすことができる．ところで，FM 波や AM 波は，無変調波と比べて広帯域を占有することが知られている．すなわち，歌声は FM や AM の効果によってコード楽器音と比較すると広帯域を占有する．この性質を周波数軸方向の滑らかさという観点から捉えなおすと，歌声は帯域幅が広いため周波数軸方向への滑らかであり，コード楽器音は帯域幅が狭いため周波数軸方向へ滑らかではない，と言い替えることができる．

表 1 歌声とコード楽器音のスペクトログラムの時間軸方向と周波数軸方向への滑らかさ
Table 1 Smoothness of spectrogram of singing voice and chordal instruments in direction of time axis and frequency axis.

	時間軸方向の滑らかさ	周波数軸方向の滑らかさ
歌声	滑らかでない	滑らか
コード楽器音	滑らか	滑らかでない

以上の性質をまとめると，歌声とコード楽器音のスペクトログラムの各軸方向への滑らかさは，大まかに**表 1** のようになる．歌声は周波数軸方向により滑らかであり，コード楽器音は時間軸方向により滑らかである．この性質に着目すると，滑らかさの方向が異なる成分を分離するような手法によって，両者を分離することができることができると考えられる．そのような処理は，我々の研究室でこれまでに開発した調波打楽器音分離を応用することで可能である．

3. 時間軸方向に滑らかな成分と周波数軸方向に滑らかな成分の分離手法

3.1 調波打楽器音分離 (HPSS)

音楽中のある成分が，周波数軸方向と時間軸方向のいずれの方向により滑らかであるかに着目してそれぞれの成分に分離する手法として，我々の研究室ではこれまでに，調波打楽器音分離 (HPSS, Harmonic/Percussive Sound Separation)[5)–8)] という手法を提案している．この手法では，ギターなどの調波楽器音（H 成分）が定常的で時間軸方向に滑らか，ドラムなどの打楽器音（P 成分）が非定常的で周波数軸方向に滑らかであることに着目して，両者を分離することを考えている．具体的には，3 つの指針

- 指針 1: H 成分は，スペクトログラム上において，時間軸方向に滑らかである．
- 指針 2: P 成分は，スペクトログラム上において，周波数軸方向に滑らかである．
- 指針 3: H 成分と P 成分を足し合わせると，入力信号と等しくなる．

に基づいて目的関数を定義し，それを最適化することで H 成分と P 成分を求める．文献 7) では，指針 1 から 3 のそれぞれの指針に関するコストの和を最小化することで H 成分と P 成分を高音質で分離しているが，分離された信号は指針 3 を厳密には満たさないため，本稿の 4 節で検討するような多重処理を行った時に，より多くの誤差が生じる可能性がある．そこで，本稿では指針 3 が厳密に成立するような拘束を与えた HPSS を新たに定式化する．

3.2 HPSS の目的関数と拘束条件

以下，入力信号，H 成分，P 成分の複素スペクトログラム（短時間フーリエ変換）をそれぞれ $W_{t,k}, H_{t,k}, P_{t,k}$ と表記する．なお，t は時間添え字，k は周波数添え字で，いずれも整数値をとり，$0 \le t \le T, 0 \le k \le K$ とする．

目的関数は，前節で述べた指針のうちの 1 および 2 を反映し，

$$J_\gamma(\{H_{t,k}\},\{P_{t,k}\}) = \sum_{t=0}^{T}\sum_{k=0}^{K}\left\{w_{\mathrm{H}}(|H_{t+1,k}|^\gamma - |H_{t,k}|^\gamma)^2 + w_{\mathrm{P}}(|P_{t,k+1}|^\gamma - |P_{t,k}|^\gamma)^2\right\} \quad (1)$$

と定義する．この目的関数の第 1 項が指針 1 に，第 2 項が指針 2 に対応している．なおスペクトログラムの絶対値を γ 乗しているのは，物理量を感覚量に換算するときに冪乗がよい近似を与えるという知見に基づくものであり，音量の場合，帯域にもよるが，感覚量は音圧のおよそ 0.6 乗に比例する[10] とされる．また，$w_{\mathrm{H}}, w_{\mathrm{P}}$ はそれぞれの項への重み係数である．なお HPSS においては，H 成分と P 成分を同時に最適化することが重要であるため，$w_{\mathrm{H}}, w_{\mathrm{P}}$ のどちらか一方に極端に傾斜をかけるのではなく，両者の違いはごくわずかである必要がある．

さらに，指針 3 を以下のような拘束条件として課す．

$$\text{for}\forall t,k,\quad W_{t,k} = H_{t,k} + P_{t,k}. \quad (2)$$

ここで，$W_{t,k}$ などは複素数であるのに対し，目的関数は $W_{t,k}$ の絶対値を含んだ非正則な関数である．このため，各変数を複素数のままで扱うことによる利点は小さい．そこで，拘束条件を絶対値と位相に分解し，代わりに

$$|W_{t,k}| = |H_{t,k}| + |P_{t,k}| \quad (3)$$

$$\angle W_{t,k} = \angle H_{t,k} = \angle P_{t,k} \quad (4)$$

を拘束条件とする．すなわち，位相は固定したまま，絶対値のみの拘束条件とみなすことで，問題を扱いやすくする．

以上の式 (1), (3) が，HPSS の目的関数と拘束条件である．

3.3 HPSS の最適化問題の実際的な解法

ここで，目的関数は $|H_{t,k}|^\gamma (\gamma \approx 0.6)$ などの関数であるのに対し，拘束条件は $|H_{t,k}|$ などに関するものになっていることに着目し，計算を簡単にするため以下 $\gamma = 0.5$ とする．これによって，両者とも $|H_{t,k}|^{0.5}$ などに関する 2 次式となる．

ここで，拘束条件を消去するために未定乗数 $\lambda_{t,k}$ を導入し，ラグランジュ関数

$$L(\{|H_{t,k}|^{0.5}\},\{|P_{t,k}|^{0.5}\},\{\lambda_{t,k}\}) = J_{\gamma=0.5}(\{H_{t,k}\},\{P_{t,k}\}) - \sum_{t=0}^{T}\sum_{k=0}^{K}\lambda_{t,k}(|W_{t,k}| - |H_{t,k}| - |P_{t,k}|) \quad (5)$$

を定義する．このラグランジュ関数は，各パラメータに関する偏微分が，解の周りでいずれも 0 にならなければならない．すなわち，for$\forall t,k$,

$$\frac{\partial L}{\partial(|H_{t,k}|^{0.5})} = (2w_{\mathrm{H}} - \lambda_{t,k})|H_{t,k}|^{0.5} - w_{\mathrm{H}}(|H_{t+1,k}|^{0.5} + |H_{t-1,k}|^{0.5}) = 0 \quad (6)$$

$$\frac{\partial L}{\partial(|P_{t,k}|^{0.5})} = (2w_{\mathrm{P}} - \lambda_{t,k})|P_{t,k}|^{0.5} - w_{\mathrm{P}}(|P_{t,k+1}|^{0.5} + |P_{t,k-1}|^{0.5}) = 0 \quad (7)$$

$$\frac{\partial L}{\partial \lambda_{t,k}} = (|W_{t,k}|^{0.5})^2 - (|H_{t,k}|^{0.5})^2 - (|P_{t,k}|^{0.5})^2 = 0. \quad (8)$$

いま，未定乗数 $\lambda_{t,k}$ を消去するため，式 (6), (7) を式 (8) へ代入すると，未定乗数 $\lambda_{t,k}$ に関する 4 次方程式が得られる．ここで $w_{\mathrm{H}} \approx w_{\mathrm{P}} \approx 1$ と仮定すると，式 (8) は $\lambda_{t,k}$ に関する 2 次方程式に近似できて，$\lambda_{t,k}$ は

$$\lambda_{t,k} \approx 2 - \frac{\sqrt{w_{\mathrm{H}}^2(|H_{t+1,k}|^{0.5} + |H_{t-1,k}|^{0.5})^2 + w_{\mathrm{P}}^2(|P_{t,k+1}|^{0.5} + |H_{t,k-1}|^{0.5})^2}}{|W_{t,k}|^{0.5}}. \quad (9)$$

これを式 (6),(7) へ代入すると，$2TK$ 元の連立方程式 for $\forall t,k$,

$$|H_{t,k}|^{0.5} = \frac{w_{\mathrm{H}}\left(|H_{t+1,k}|^{0.5} + |H_{t-1,k}|^{0.5}\right)|W_{t,k}|^{0.5}}{\sqrt{w_{\mathrm{H}}^2\left(|H_{t+1,k}|^{0.5} + |H_{t-1,k}|^{0.5}\right)^2 + w_{\mathrm{P}}^2\left(|P_{t,k+1}|^{0.5} + |P_{t,k-1}|^{0.5}\right)^2}} \quad (10)$$

$$|P_{t,k}|^{0.5} = \frac{w_{\mathrm{P}}\left(|P_{t,k+1}|^{0.5} + |P_{t,k-1}|^{0.5}\right)|W_{t,k}|^{0.5}}{\sqrt{w_{\mathrm{H}}^2\left(|H_{t+1,k}|^{0.5} + |H_{t-1,k}|^{0.5}\right)^2 + w_{\mathrm{P}}^2\left(|P_{t,k+1}|^{0.5} + |P_{t,k-1}|^{0.5}\right)^2}} \quad (11)$$

が得られる．この連立方程式は非常に大規模であり陽に解くことは困難だが，式 (10), (11) の各右辺を評価した結果を，各左辺の近似値として各左辺へ逐次的に代入するような操作を繰り返すことによって，この問題の近似解を得ることができる．

これにより得られた解 $\{|H_{t,k}|^{0.5}\},\{|P_{t,k}|^{0.5}\}$ と位相情報 (4) から，複素スペクトログラ

表 2 コード楽器音，歌声，打楽器の音のそれぞれを，短いフレームと長いフレームのスペクトログラム上で HPSS によって分離したときに，H 成分と P 成分のどちらに分離されやすいかの傾向．
Table 2 Tendency into which the sound tends to be separated by HPSS, H component or P component. The sounds are chordal instruments, singing voice, and percussion. The lengths of analyzing frames of are about 30[ms] and 200[ms].

	短いフレームを用いた従来法 (30[ms] 程度)	長いフレーム (200[ms] 程度)
コード楽器音	H 成分へ分離	H 成分へ分離
歌声	H 成分へ分離	P 成分へ分離
打楽器	P 成分へ分離	P 成分へ分離

ム $\{H_{t,k}\}, \{P_{t,k}\}$ を復元し，逆短時間フーリエ変換により波形情報を求めることで，所望の H 成分（調波的成分）と P 成分（打楽器的成分）が得られる．各要素 $\{|H_{t,k}|^{0.5}\}, \{|P_{t,k}|^{0.5}\}$ に関する更新の回数は，経験的に，5 回程度である程度の精度の近似解が得られ，50 回程度である値へほぼ収束する．なおこの反復計算は，分析ブロックを用いることでオンラインで実行することができる[7]．

4. 多重 HPSS 法

4.1 長い窓関数を用いた HPSS による歌声とコード楽器音の分離

前節で述べた HPSS は，本来はギターなどコード楽器音を H 成分に，打楽器などを P 成分に分離する手法だが，これはスペクトログラム上での滑らかさの方向に着目した手法であるため，同様な方法によってコード楽器音と歌声とを分離することができると考えられる．

ただし，従来の HPSS をそのまま適用するだけでは，歌声成分はコード楽器音と同様に H 成分へと分離されてしまい，両者を分離することができない．これは，通常のスペクトログラム分析で用いるような 30[ms] 程度の時間幅のフレーム（窓関数）で分析した場合，たかだか 5–8Hz 程度の歌声のスペクトルゆらぎの効果よりも，短い窓関数を用いたことで周波数分解能が粗くなることの影響の方が大きくなるためである．すなわち，歌声とコード楽器音の違いがスペクトログラムの滑らかさの方向の違いとして現れなくなるため，スペクトログラム上で近接した数 bin のみの情報からこれらを識別することができない．

そこで，歌声とコード楽器音を HPSS によって識別するためには，フレームの時間幅を従来の HPSS よりも長く取り，歌声の非定常性が単一フレーム内で観測できるようにすればよいと考えられる．そのために必要な窓関数の長さはヴィブラートの 1 周期程度以上と仮定すれば，歌声のヴィブラートが 5–8Hz 程度であることから，フレーム長はおよそ 200[ms] 以上と想定することができる．

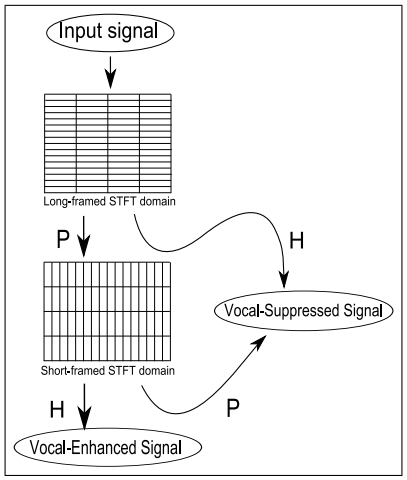

図 1 直列多重 HPSS 法の概念図
Fig. 1 Diagram of serial multi-stage HPSS.

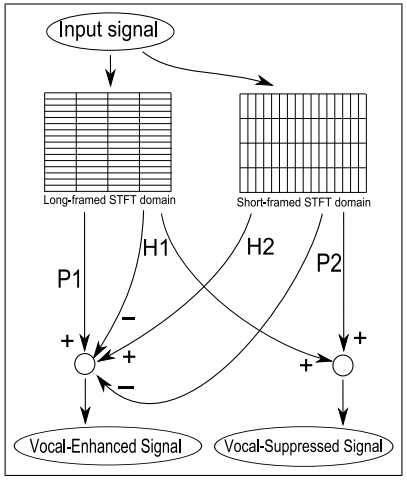

図 2 並列多重 HPSS 法の概念図
Fig. 2 Diagram of parallel multi-stage HPSS.

以上の議論をまとめると，スペクトログラムを計算するときに用いるフレームの時間幅に依存して，コード楽器音，歌声，打楽器のそれぞれは**表 2** のような性質を示すと考えることができる．

4.2 多重 HPSS 法による歌声と伴奏（コード楽器音 + 打楽器音）の分離

コード音，歌声，打楽器音のそれぞれの音が，H 成分，P 成分のいずれに分離されやすいかは，表 2 のように，スペクトログラムのフレームの時間幅に依存する．また，HPSS の重み係数 $w_{\mathrm{H}}, w_{\mathrm{P}}$ の微妙な差も同様に分離性能に影響を与え，例えば w_{P} を w_{H} より若干小さく $(w_{\mathrm{H}}, w_{\mathrm{P}}) = (1.00, 0.95)$ などとすると，P 成分に分離されるための条件が厳しくなるため，より P 的な成分のみが P に分離されるようになるという性質がある．

これらを利用すると，スペクトログラムを長短それぞれのフレームで計算し，それぞれの領域上で適当な重み係数を用いて HPSS を行うことによって，これらの混合信号の各成分を分離することができると考えられる．このような分離信号が得られれば，それらを適当に再合成することで，特定の成分を強調した信号，すなわち歌声強調信号や伴奏強調信号が得られる．

このような方法の一つとして考えられるのが，**図 1** のように，入力された音楽信号に対

して HPSS を段階的に 2 回適用することで，コード楽器音，歌声，打楽器音を分離する方法である[11]．この方法では，まず，長いフレームのスペクトログラム上で，P に重点をおいた重み係数を用いた HPSS を適用することで，P 成分（打楽器的成分 + 歌声的成分）と H 成分（コード楽器音的成分）を分離する．次に，前の処理で得られた H 成分を，短いフレームのスペクトログラム上で，H に重点をおいた重み係数を用いた HPSS を適用することで，P 成分（打楽器的成分）と H 成分（歌声的成分）とに分離すれば，ここで得られた H 成分を「歌声」とすることができる．また，その際に副産物として得られる打楽器的成分とコード楽器音的成分を用いて，それらを足し合わせた信号を「伴奏」とすることもできる．この方法は直列多重 HPSS 法と呼ぶこととする．

もう一つの方法として考えられるのが，**図 2** のように，入力された音楽信号に対し 2 種類のスペクトログラム上で HPSS を同時に実行する方法である．2 種類のうちの一方では P_1 成分（打楽器的成分 + 歌声的成分）と H_1 成分（コード楽器音成分）を分離し，もう一方では P_2 成分（打楽器的成分）と H_2 成分（歌声的成分 + コード楽器音的成分）の分離を行うことで，それぞれの出力を用いて「歌声」と「伴奏」を

$$\text{「歌声」} = (\mathrm{P}_1 + \mathrm{H}_2 - \mathrm{H}_1 - \mathrm{P}_2)/2 \tag{12}$$

$$\text{「伴奏」} = \mathrm{H}_1 + \mathrm{P}_2 \tag{13}$$

として求めることができる．この方法は並列多重 HPSS 法と呼ぶこととする．

5. 実音楽データを用いた実験

5.1 実験条件

多重 HPSS 法の効果を確認するため，実際の音楽音響信号を用いて，歌声強調/抑圧の実験を行った．実験に用いたのは，RWC 研究用音楽データベース[3] より，ポップス，ジャズ，および著作権切れ楽曲である．これらはいずれもモノラル信号に変換し，16kHz にリサンプリングして用いた．これらの楽曲に対し，直列多重 HPSS 法，並列多重 HPSS 法のそれぞれを適用し，得られた信号を聴くことで，この手法の定性的な性質を調べた．

短時間フーリエ変換に使った窓関数は，順変換，逆変換ともにサイン窓（ハニング窓の平方根）で，窓関数は半分ずつオーバーラップさせながらスペクトログラムを求めた．片方のスペクトログラムでは，フレーム長は 256[ms](4096 点)，重み係数は $(w_{\mathrm{H}}, w_{\mathrm{P}}) = (1.00, 0.95)$ とした．もう一方のスペクトログラムでは，フレーム長は 32[ms](512 点)，重み係数は $(w_{\mathrm{H}}, w_{\mathrm{P}}) = (0.95, 1.00)$ とした．いずれも HPSS の更新式 (10), (11) の反復回数は 30 回とした．また，重み係数 $w_{\mathrm{H}}, w_{\mathrm{P}}$ の値を相互に入れ替えた場合についても同様な実験を行った．

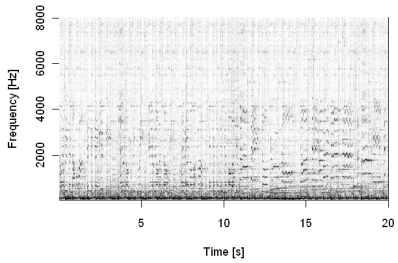

図 3 RWC データベースより，ポピュラー楽曲（RWC-MDB-P-2001 No. 96）のスペクトログラム
Fig. 3 Spectrogram of a popular music from RWC music database.

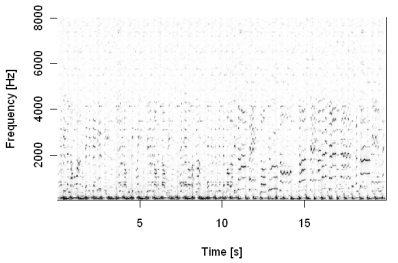

図 4 図 3 の楽曲に直列多重 HPSS 法を適用したときのスペクトログラム
Fig. 4 Output spectrogram of serial multi-stage HPSS.

5.2 実験結果

様々な楽曲に多重 HPSS 法を適用した結果観察された定性的な性質として，全体的な傾向として，以下のような音が「歌声」に分離されやすいことが観察された．

- 歌声で，特にヴィブラートがかかった音や，旋律的な音
- ピアノなどの音の立ち上がりの瞬間
- ヴァイオリンやトランペットなど，ヴィブラートなどがかかりやすい音
- ベースラインやバスドラムなど低音域の音
- 対旋律など，伴奏の中でも比較的旋律的な音

一方，以下のような音が「伴奏」に分離されやすいことが観察された．

- 歌声の，立ち上がりの瞬間や，平坦で音価の長い比較的コード楽器音的な音
- ギターやピアノなどの，立ち上がり以外の部分
- スネアドラムなどの打楽器音

直列多重 HPSS 法と並列多重 HPSS 法の結果を比較したとき，前者の方が後者よりも「歌声」と「伴奏」のそれぞれが排他的（「歌声」では伴奏があまり聞こえず,「伴奏」では歌声はあまり聞こえない）に分離される傾向にあった．また，H と P への重み係数を逆にした場合と比較すると，通常の場合では「歌声」は伴奏に対して排他的で「伴奏」にはやや歌声が混ざりやすい傾向があったのに対し，重み係数を逆にした場合ではその逆の性質を示し，「伴奏」は歌声に対して排他的で「歌声」にはやや伴奏が混ざりやすい傾向があった．

最後に，実験に用いた楽曲のうちの 1 曲の入力信号，および直列多重 HPSS 法により求めた「歌声」強調信号のスペクトログラムを**図 3**, **図 4** に示す．図 3 に見られる縦の筋（打楽器音），横の筋（コード楽器音）が，図 4 では抑圧されていることが見られる．

6. 考　　察

多重 HPSS 法を実音楽信号へ適用する実験によって，提案法によって歌声に相当する成分の大部分は強調され，伴奏に相当する成分の大部分は抑圧されることが確認された．また同時に，楽器音でも，音の立ち上がりの部分や，微細なゆらぎのある音や，旋律的な動きをする場合は,「歌声」として強調されやすい傾向にあることが確認された．なお，これらの性質は，始めに我々が歌声の性質として想定していた性質でもあり，このことは，本手法の一定の有効性を示すものである．すなわち，本手法の適用後，あるいは適用前に，楽器音にはないような歌声に特有な条件を用いた処理を施すことで，より高精度の歌声強調も容易に実現することができると考えられる．

H, P の重みに関しては，2 種類の実験によって，歌声における伴奏の排他性を重視するか，伴奏における歌声の排他性を重視するかをある程度制御できることが確認された．これらの中からどれを利用するかは応用次第であり，例えば後者で求めた「伴奏」成分は，カラオケとして使うときに有効と考えられる．

直列多重 HPSS 法と並列多重 HPSS 法とを比較したとき，時間遅延という観点からすると，並列版では，この遅延が 1 段階の HPSS と同程度であるのに対し，直列版の場合は，2 段階の HPSS によってその遅延が拡大してしまうという欠点がある．一方音質という観点では，直列版は並列版よりも音質が優れている場合が多い．本手法を応用するときに，どちらの方法を採用するかは遅延の短さと音質のどちらを優先するか次第である．

7. 結　　論

本稿では，音楽中の歌声を強調/抑圧するために，歌声のヴィブラートや旋律的な動きなどといったスペクトルの時間変動が，時間幅の短さや帯域幅の広さとなってスペクトグラム上に現れるという性質に着目した．この性質を検出するために，スペクトルの時間軸方向，周波数軸方向のそれぞれの方向への滑らかさを基準に楽器音を分離する手法である調波打楽器音分離 (HPSS) を 2 回用いる方法を提案し，実際の音楽音響信号にこの手法を適用した結果，ある程度音質を制御しながら，歌声強調信号/抑圧信号が得られることが確認された．また本手法によって，楽器音でも，音の立ち上がりや，微細なゆらぎのある音に関しても,「歌声」として強調されやすい傾向にあることが確認された．

本稿では定量的な分離性能を評価するまでには至っていないが，今後，定量評価を行える環境を構築することによって，歌声強調に最適なフレーム長や重み係数の値を調べることを検討している．また，本研究の成果として得られた歌声強調手法に，後処理として基本周波数推定手法などを用いることで，メロディラインを認識するなどの応用も検討している．

参 考 文 献

1) Downie, J.S.: The music information retrieval evaluation exchange(2005-2007): A window into music information retrieval research, *Acousti. Sci. & Tech.*, Vol. 29, No.4, pp.247–255 (2008).
2) 藤原弘将，北原鉄朗，後藤真孝，駒谷和範，尾形哲也，奥乃　博：伴奏音抑制と高信頼度フレーム選択に基づく楽曲の歌手名同定手法，情報処理学会論文誌， Vol.47, No.6, pp.1831–1843 (2006).
3) 後藤真孝，橋口博樹，西村拓一，岡　隆一：RWC 研究用音楽データベース: 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース，情報処理学会論文誌， Vol.45, No.3, pp.728–738 (2004).
4) Li, Y. and Wang, D.L.: Separation of Singing Voice From Music Accompaniment for Monaural Recordings, *IEEE Transactions on Audio, Speech, and Language Processing*, Vol.15, No.4 (2007).
5) 宮本賢一，立薗真理，ルルージョナトン，亀岡弘和，小野順貴，嵯峨山茂樹：スペクトログラム 2 次元フィルタによる調波音・打楽器音の分離，日本音響学会講演論文集（秋），pp.825–826 (2007).
6) 宮本賢一，亀岡弘和，小野順貴，嵯峨山茂樹：スペクトログラムの滑らかさの異方性に基づいた調波音・打楽器音の分離，日本音響学会講演論文集（春），pp.903–904 (2008).
7) Ono, N., Miyamoto, K., Kameoka, H. and Sagayama, S.: A Real-Time Equalizer of Harmonic and Percussive Components in Music Signals, *Proceedings of ISMIR*, pp.139–144 (2008).
8) Ono, N., Miyamoto, K., Le Roux, J., Kameoka, H. and Sagayama, S.: Separation of a Monaraul Audio Signal into Harmonic/Percussive Components by Complementary Diffusion on Spectrogram, *Proceedings of EUSIPCO* (2008).
9) Saitou, T., Unoki, M. and Akagi, M.: Development of an F0 control model based on F0 dynamics characteristics for singing-voice synthesis, *Speech Communication*, Vol.5, pp.267–277 (2005).
10) Stevens, S.S.: The Measurement of Loudness, *The Journal of the Acoustic Society of America*, Vol.27, No.5, pp.815–829 (1955).
11) 橘　秀幸，小野順貴，嵯峨山茂樹：多重 HPSS 法によるモノラル音楽音響信号に対するボーカル抑圧，日本音響学会講演論文集（春），pp.852–853 (2009).