スペクトログラムの滑らかさの異方性に基づく 調波音・打楽器音分離の各手法の性能比較*

◎橘秀幸 (東大・情報理工), 亀岡弘和 (東大・情報理工/NTT),
小野順貴 (NII), 嵯峨山茂樹 (東大・情報理工)

1 はじめに

音楽信号には，ギターのような調波楽器音と，スネアドラムのような打楽器音などのように，性質の大きく異なる複数の楽器音が混合している．本稿ではこれらを分離する問題を扱う．このような技術は音楽情報処理のための前処理としても有効であり，また，調波楽器音と打楽器音のイコライザのような新しい音楽鑑賞の形態への応用も考えることができることから，調波音と打楽器音の分離手法はこれまでにいくつかの研究が行われている [1].

そのような手法のひとつとして，著者らはこれまでに特定の楽器に関する事前知識などを使わずに調波音と打楽器音を分離する手法，調波音・打楽器音分離 (HPSS) を提案している．HPSS は，調波音のスペクトログラムは時間方向に滑らか（水平成分）である一方，打楽器音のスペクトログラムは周波数方向に滑らか（垂直成分）であるという，スペクトログラムの滑らかさの異方性に着目し，水平な成分と垂直な成分とを分離することによって信号を分離する手法である．HPSS は音楽信号処理の様々な手法のための前処理として有効であることが確認されており [2], 音楽音響信号の和音認識の性能改善 [3]，楽曲のリズムマップの作成 [4]，楽曲中のメロディ推定の前処理 [5, 6] など様々な応用に利用されている．

ところで，スペクトログラムの滑らかさを評価する規準や，パラメータ設定等にはいくつかの選択肢が考えられ，それらの選択肢に応じて複数の手法を提案している [7, 8, 9, 10, 11]. HPSS は音楽情報処理の前処理等として重要であることから，各手法のうちどの手法が最も性能が高いかを調べることは，応用の面でも重要であると考えられる．本稿では，これらの各手法の関係を整理し，また各手法の性能を同一の音楽音響信号と評価基準を用いて評価して各手法の比較を行った．

2 調波・打楽器音分離のフレームワーク

本稿では以下の記号を用いる．$x(t)$ は信号の波形情報とする．$\boldsymbol{X}^{(L)} = (X_{\tau,k,L})_{1\le t\le T, 1\le k\le K}$ は，$x(t)$ を短時間フーリエ変換することより得られるパワースペクトログラムである．ここで，τ は時間，k は周波数のインデックスで，T, K はそれぞれの最大値である．また，L は短時間フーリエ変換のフレーム長であるが，以後省略して表記する．なお，本稿では，短時間フーリエ変換のフレームシフトは $L/2$，窓関数はハニング窓の平方根に固定する．

HPSS では，入力信号 $w(t)$ のパワースペクトログラム $\boldsymbol{W}$ を，Fig. 1 のように，時間方向になめらかなパワースペクトログラム $\boldsymbol{H}$ と周波数方向になめらかなパワースペクトログラム $\boldsymbol{P}$ とに分離し，それに位相情報を与えて逆短時間フーリエ変換することにより，調波的成分 $h(t)$ と打楽器的成分 $p(t)$ とに分離する．すなわち，HPSS は，パワースペクトログラム $\boldsymbol{W}$ が与えられたときに下記のような性質を満たすパワースペクトログラム $\boldsymbol{H}, \boldsymbol{P}$ を求める問題として定式化することができる．

1. $\boldsymbol{H}$ は時間方向に，$\boldsymbol{P}$ は時間方向にそれぞれ滑らか

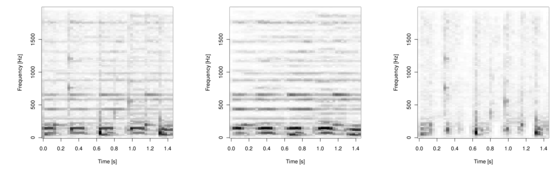

Fig. 1 HPSS による分離の例. 左：原信号 $\boldsymbol{W}$，中：分離された調波的成分 $\boldsymbol{H}$，右：同，打楽器的成分 $\boldsymbol{P}$

2. 推定された信号の和は原信号にほぼ一致する．すなわち $h(t) + p(t) \approx w(t)$.
3. 推定されたパワースペクトログラムが非負．すなわち $\forall(t,k), H_{\tau,k} \ge 0, P_{\tau,k} \ge 0$.

以下，1 と 2 について詳しく述べる．

2.1 スペクトログラムの滑らかさと異方性

スペクトログラムの時間方向，周波数方向の滑らかさの規準として，著者らはこれまでに，各方向への差分の L^2 ノルムを利用することを検討している．すなわち，スペクトログラム $\boldsymbol{X}$ が時間方向に滑らかである場合，$\boldsymbol{X}$ の時間方向差分の絶対値が小さいと考えられることから，$\boldsymbol{X}$ の時間方向の滑らかさを

$$\Omega_{\mathrm{H}}(\boldsymbol{X}, \gamma) = \sum_{\tau=0}^{T-1} \sum_{k=0}^{K} ({H_{\tau+1,k}}^{\gamma} - {H_{\tau,k}}^{\gamma})^2 \qquad (1)$$

により評価することができる．スペクトログラム $\boldsymbol{X}$ が時間方向に滑らかであるとき，この値は小さくなる．同様に，周波数方向への滑らかさは，

$$\Omega_{\mathrm{P}}(\boldsymbol{X}, \gamma) = \sum_{\tau=0}^{T} \sum_{k=0}^{K-1} ({P_{\tau,k+1}}^{\gamma} - {P_{\tau,k}}^{\gamma})^2 \qquad (2)$$

により評価することができる．なお，ここで γ 乗は，信号のパワーを人間の聴覚のスケールに近似的に変換するための指数であり，0.3 程度の値が望ましいと考えられる．

2.2 分離結果の和と原信号の一致度

HPSS では，単に時間方向・周波数方向に滑らかな成分を求めるのではなく，両成分の分離を行うことが目的であるため，分離した信号の和がもとの信号に戻るように $\boldsymbol{H}, \boldsymbol{P}$ を求める必要がある．すなわち，分離した調波成分と打楽器成分の和が，原信号にできるだけ近くなるような拘束が必要となる．そのための方法としては，それぞれ拘束の強い順に以下の 3 つを考えることができる．

1. ${W_{\tau,k}}^{\gamma} = {H_{\tau,k}}^{\gamma} + {P_{\tau,k}}^{\gamma}$ の拘束条件のもとで，各滑らかさ規準の重み付き和 $\Omega_{\mathrm{H}}(\boldsymbol{H}, \gamma) + \kappa\Omega_{\mathrm{P}}(\boldsymbol{P}, \gamma)$ を最小化する．
2. 原信号と調波・打楽器各成分の和の乖離度 $d(\boldsymbol{W}, \boldsymbol{H} + \boldsymbol{P})$ を，$\Omega_{\mathrm{H}}(\boldsymbol{H}, \gamma) + \kappa\Omega_{\mathrm{P}}(\boldsymbol{P}, \gamma)$ と同時に小さくするような最適化を行う．

---

* Hideyuki Tachibana[1], Hirokazu Kameoka[1,2], Nobutaka Ono[3], Shigeki Sagayama[1], (1. Graduate School of Information Science and Technology, The University of Tokyo, 2. NTT, 3. National Institute of Informatics,) "Performance Comparison of Harmonic/Percussive Sound Separation Techniques Based on Anisotropic Smoothness of Spectrogram"

3. この条件を陽には扱わずに $\boldsymbol{H}, \boldsymbol{P}$ を大まかに推定した後，それらを用いて時間周波数マスク（典型的にはウィーナーマスクおよびバイナリマスク）を設計し，改めて分離を行う．

なお手法 1 に関しては，$\boldsymbol{H}, \boldsymbol{P}$ を逆短時間フーリエ変換するときの位相情報に原信号の位相情報が同一と仮定したとき，$h(t)+p(t)=w(t)$ は ${H_{\tau,k}}^{1/2}+{P_{\tau,k}}^{1/2}={W_{\tau,k}}^{1/2}$ は等価である．また，手法 3 は手法 1,2 との両立が可能であり，手法 1,2 の結果をマスク設計に利用することも考えられる．

3 調波打楽器音分離の具体的な実現方法

3.1 拘束条件 $h(t)+p(t)=w(t)$ 下での滑らかさ規準最適化に基づく手法 1 (HPSS-HM1)

HPSS の具体的な実現方法のひとつとしては，2.1 節で述べた滑らかさ規準の重みつき和

$$J_{\gamma,\kappa_1}(\boldsymbol{H},\boldsymbol{P})=\Omega_{\mathrm{H}}(\boldsymbol{H},\gamma)+\kappa_1\Omega_{\mathrm{P}}(\boldsymbol{P},\gamma) \tag{3}$$

を，2.2 節の 1 で述べたような拘束条件

$${H_{\tau,k}}^{\gamma}+{P_{\tau,k}}^{\gamma}-{W_{\tau,k}}^{\gamma}=0 \tag{4}$$

のもとで最小化するというアプローチを考えることができる [9]．ただし κ_1 は重み係数である．この目的関数 $J_{\gamma,\kappa_1}(\boldsymbol{H},\boldsymbol{P})$ は以下のような手続きを反復することにより最小化することができる．

$$\begin{aligned}{H_{\tau,k}}^{\gamma} &\leftarrow \min(\max(\beta,0),{W_{\tau,k}}^{\gamma}), &(5)\\ {P_{\tau,k}}^{\gamma} &\leftarrow {W_{\tau,k}}^{\gamma}-{H_{\tau,k}}^{\gamma}, &(6)\end{aligned}$$

where

$$\begin{aligned}\beta &= {H_{\tau,k}}^{\gamma}\\ &+ \frac{1}{4(1+\alpha)}({H_{\tau+1,k}}^{\gamma}-2{H_{\tau,k}}^{\gamma}+{H_{\tau-1,k}}^{\gamma})\\ &- \frac{\alpha}{4(1+\alpha)}({P_{\tau,k+1}}^{\gamma}-2{P_{\tau,k}}^{\gamma}+{P_{\tau,k-1}}^{\gamma}). \quad (7)\end{aligned}$$

なお γ に関しては，前述のように $h(t)+p(t)=w(t)$ が厳密に成立するためには $\gamma=0.5$ である必要があるが，拘束条件の厳密性よりも滑らかさ規準を聴覚特性に近づけることを優先し，$\gamma=0.3$ などと設定することも可能である．

3.2 拘束条件 $h(t)+p(t)=w(t)$ 下での滑らかさ規準最適化に基づく手法 2(HPSS-HM2)

滑らかさ規準における γ の値は 0.3 に近い値であることが望ましいと考えられるが，一方で $\gamma=0.3$ とすると，拘束条件 $h(t)+p(t)=w(t)$ は厳密には満たされない．そこで，滑らかさ規準に関する γ のみを 0.3 に近い値にする一方，拘束条件 (4) に関しては $\gamma=0.5$ を用いることを考える．すなわち，拘束条件を (4) に $\gamma=1/2$ を代入したものを用い，目的関数としては，計算を簡単にするため 0.3 の近似値として $\gamma=1/4$ を用いたもの $J_{1/4,\kappa_2}(\boldsymbol{H},\boldsymbol{P})$ を用いる [11]．

これらの目的関数および拘束条件に関してラグランジュ未定乗数法を適用し，簡単のため $\kappa_2\approx 1$ と仮定すると，以下のような関係式が得られる．

$$\begin{aligned}{H_{\tau,k}}^{1/2} &= \alpha_{\tau,k}{W_{\tau,k}}^{1/2}/(\alpha_{\tau,k}+\beta_{\tau,k}), &(8)\\ {P_{\tau,k}}^{1/2} &= \beta_{\tau,k}{W_{\tau,k}}^{1/2}/(\alpha_{\tau,k}+\beta_{\tau,k}), &(9)\end{aligned}$$

where

$$\begin{aligned}\alpha_{\tau,k} &= ({H_{\tau+1,k}}^{1/4}+{H_{\tau-1,k}}^{1/4})^2, &(10)\\ \beta_{\tau,k} &= \kappa_2^2({P_{\tau,k+1}}^{1/4}+{P_{\tau,k-1}}^{1/4})^2. &(11)\end{aligned}$$

この $2TK$ 元連立方程式を直接解くことは困難だが，これらを更新式と見なし代入を繰り返すことによって適当な $\boldsymbol{H}, \boldsymbol{P}$ を求めることができる．

3.3 $\boldsymbol{H}+\boldsymbol{P}$ と $\boldsymbol{W}$ の乖離度を I ダイバージェンスにより評価する手法 (HPSS-Idiv)

ここまでに述べたような $\boldsymbol{W}=\boldsymbol{H}+\boldsymbol{P}$ を拘束条件として取り扱うアプローチの他に，2.2 節の 2 で述べたような，$J_{\gamma,\kappa}(\boldsymbol{H},\boldsymbol{P})$ と同時に信号同士の乖離度 $d(\boldsymbol{W},\boldsymbol{H}+\boldsymbol{P})$ を最小化するアプローチも考えることができる．この場合，関数 $d(\cdot,\cdot)$ の具体的な形が問題となるが，パワースペクトル同士の乖離度に関する指標として，特に I ダイバージェンスが有効であることが知られていることから，これを利用することが有効と考えられる．I ダイバージェンスは，パワースペクトログラム $\boldsymbol{X}$ と $\boldsymbol{Y}$ に関して

$$D_I(\boldsymbol{X}|\boldsymbol{Y})=\sum_{\tau,k}X_{\tau,k}\log\frac{X_{\tau,k}}{Y_{\tau,k}}-(X_{\tau,k}-Y_{\tau,k}) \tag{12}$$

により定義される量であり，この値が小さいほど両者が近いことを示す．このような観点に基づき，文献 [10] では目的関数を

$$\frac{\Omega_{\mathrm{H}}(\boldsymbol{H},\gamma)}{\sigma_{\mathrm{H}}^2}+\frac{\Omega_{\mathrm{P}}(\boldsymbol{P},\gamma)}{\sigma_{\mathrm{P}}^2}+D_I(\boldsymbol{W}|\boldsymbol{H}+\boldsymbol{P}) \tag{13}$$

とし，これを最小化することにより $\boldsymbol{H}, \boldsymbol{P}$ を求める手法を提案している．ただし $\sigma_{\mathrm{H}}, \sigma_{\mathrm{P}}$ は適当な定数である．ところで，目的関数の第 1 項と第 2 項は信号のパワーの 2γ 乗に比例するのに対し，第 3 項はパワーに比例するから，両者の重みが信号のスケールに依存してしまう．このようなスケールへの依存性を防ぐためには，$2\gamma=1$ すなわち $\gamma=0.5$ とする必要がある．

この目的関数は以下のような手続きを反復することにより最適化することができる．

$$\begin{aligned}H_{\tau,k} &\leftarrow \left(\frac{B_1+\sqrt{B_1^2+4A_1C_1}}{2A_1}\right)^2 &(14)\\ P_{\tau,k} &\leftarrow \left(\frac{B_2+\sqrt{B_2^2+4A_2C_2}}{2A_2}\right)^2 &(15)\\ m_{\tau,k} &\leftarrow H_{\tau,k}/(H_{\tau,k}+P_{\tau,k}) &(16)\end{aligned}$$

where

$$\begin{aligned}&A_1=2/\sigma_{\mathrm{H}}^2+2,\ A_2=2/\sigma_{\mathrm{P}}^2+2, &(17)\\ &B_1=({H_{\tau+1,k}}^{1/2}+{H_{\tau-1,k}}^{1/2})/\sigma_{\mathrm{H}}^2, &(18)\\ &B_2=({P_{\tau,k+1}}^{1/2}+{P_{\tau,k-1}}^{1/2})/\sigma_{\mathrm{P}}^2, &(19)\\ &C_1=2m_{\tau,k}W_{\tau,k},\ C_2=2(1-m_{\tau,k})W_{\tau,k}. &(20)\end{aligned}$$

この方法により得られる $\boldsymbol{H}, \boldsymbol{P}$ の和は原信号には必ずしも一致しないものの，2.2 節にて述べたように，後処理として時間周波数マスキングを適用することにより両者を厳密に一致させることも可能である．

3.4 2 次元フィルタによる時間周波数マスク設計に基づく手法 (HPSS-2DFilter)

ところで，2.2 節の 3 に述べたように，後処理として時間周波数マスクをかけることを前提とした場合，前節までに述べた各手法のような，目的関数の最適化に基づいたアプローチをとることは必ずしも必要ではなく，より簡単な計算により仮の $\boldsymbol{H}, \boldsymbol{P}$ を求めるのみでも十分である可能性があると考えられる．そのような方法のひとつとして，スペクトログラム上の 2 次元のフィルタに基づく方法が考えられる [7]．

いま仮の $\boldsymbol{H}$ として，原信号のスペクトログラム ${W_{\tau,k}}^{\gamma}$ に関し，時間方向に平滑化（フィルタ係数：α_i）することにより得られるスペクトログラム ${\hat{H_{\tau,k}}}^{\gamma}=\sum_{i=-I}^{I}\alpha_i {W_{(t+i),k}}^{\gamma}$ を考え，仮の $\boldsymbol{P}$ として原信号の

Table 1　因子とその水準

手法	因子	水準
HM1 2DFilter	γ	0.3, 0.4, 0.5
HM2 2DFilter	κ_1, c	0.25, 0.5, 0.6, 0.9 1, 1.2, 1.5, 2, 4
HM2	κ_2	$0.8 + 0.04n$ $(n = 1, \cdots, 9)$
Idiv	$\sigma_{\mathrm{H}}, \sigma_{\mathrm{P}}$	$0.1, 0.3, 0.5$
2DFilter	I, J	10, 50, 200
HM1 HM2 Idiv	更新式の 反復回数	10, 30, 300
全手法	時間 周波数 マスク	None, Wiener, Binary
全手法	L	256, 512, 1024 (16 kHz)

スペクトログラム ${W_{\tau,k}}^{\gamma}$ に関し，周波数方向に平滑化（フィルタ係数：β_i）することにより得られるスペクトログラム $\hat{P_{\tau,k}}^{\gamma} = \sum_{i=-J}^{J} \beta_i {W_{\tau,(k+j)}}^{\gamma}$ を考える．このようにして得られた $\hat{\boldsymbol{H}}, \hat{\boldsymbol{P}}$ の直接的な逆短時間フーリエ変換は必ずしも十分な音質にはならないと考えられるが，これらを利用して時間周波数マスクを設計することにより，$\boldsymbol{H} + \boldsymbol{P} = \boldsymbol{W}$ であるような $\boldsymbol{H}, \boldsymbol{P}$ を得ることができる．

4　HPSS の各手法の性能評価実験

4.1　実験計画法に基づく各手法のパラメータ決定およびそれに基づく音楽信号の分離性能評価

HPSS の各手法には変数や後処理に多数の選択肢があるが，これらのパラメータには膨大な組み合わせがあり，その全てを試みて最適な組み合わせを探すのは現実的ではない．そこで本稿では，実験計画法 [12] に基づき，これらの選択肢のうちから最適な組み合わせの探索を行った．

本実験によって決定したパラメータは，HM1 の κ_1, γ，HM2 の κ_2，Idiv の $\sigma_{\mathrm{H}}, \sigma_{\mathrm{P}}$，2DFilter の $\{\alpha_\tau\}, \{\beta_k\}$ である．なお，$\{\alpha_\tau\}, \{\beta_k\}$ に関しては自由度が大きすぎるため，$\alpha_\tau = c(0.5 + 0.5\cos(\tau\pi/I)), \beta_k = 0.5 + 0.5\cos(k\pi/J)$ とし，フィルタ長 I, J と重み c をパラメータとした．また，後処理として用いる時間周波数マスクの種類，更新式の反復回数，および短時間フーリエ変換で使用するフレーム長 L にも選択肢がある．それぞれのパラメータ（因子）のとりうる値（水準）は Table 1 の通りとした．

実験には，文献 [1] で用いられている音楽データと同じものを使用した．これらは主に MASS データベース [13] より抜粋された 6 データであり，様々なジャンルの楽曲を含んでいる．これらの楽曲はいずれも，サンプリング周波数 16 kHz のモノラル信号で，長さは 10 秒程度である．各試行での性能の評価には，(A) 全体の音質を評価する指標として，$[h(t)\ p(t)]^{\mathsf{T}}$ の SDR，(B) 調波音の歪みの小ささを優先して評価する指標として，$h(t)$ の SDR の改善値，および (C) 打楽器音の歪みの小ささを優先して評価する指標として，$p(t)$ の SDR の改善値，を用いた．ただし SDR は目的信号を $\boldsymbol{x}(t) = [x_1(t) \ldots x_n(t)]^{\mathsf{T}}$，推定信号を $\boldsymbol{y}(t) = [y_1(t) \ldots y_n(t)]^{\mathsf{T}}$ としたときに

$$\mathrm{SDR} = 10 \log_{10} \frac{E_t[\boldsymbol{x}^{\mathsf{T}}\boldsymbol{y}]^2}{E_t[\boldsymbol{x}^{\mathsf{T}}\boldsymbol{x}]E_t[\boldsymbol{y}^{\mathsf{T}}\boldsymbol{y}] - E_t[\boldsymbol{x}^{\mathsf{T}}\boldsymbol{y}]^2}, \quad (21)$$

により定義される値である．

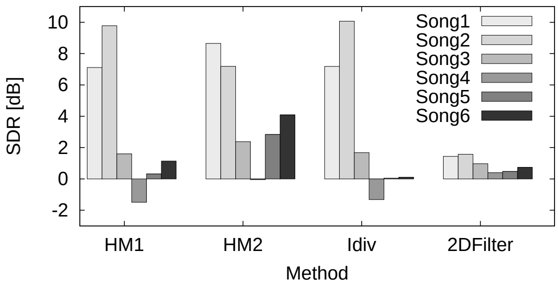

Fig. 2　各手法によって得られた各楽曲の分離信号の SDR

これらの条件で，それぞれの因子を $L_{27}(3^{13})$ 直交表に割付を行い，各音楽信号につき 27 通りのパラメータセットで実験を行った．また，その実験結果に基づいて分散分析に基づき，各パラメータを決定した．

Table 2 に本実験により決定したパラメータセットを示す．なお，括弧がついていない因子は，性能への影響に関して有意水準 1%で有意性が確認された因子であり，それぞれに関して効果が最大となった水準を採用した．また，括弧内の因子は必ずしも分離性能への影響の有意性は確認されていない因子だが，本実験において効果が最大となった水準を採用した．

以上により決定したパラメータセットのうち全体の SDR に関して最適なパラメータセット (A) を用いて 6 曲のデータを分離したときの分離性能の評価を行った．Fig. 2 は，各手法を用いて各信号を分離したときの SDR を示している．HM1, HM2, Idiv に関しては，分離性能は概ね楽曲に依存することが観察できるが，特に HM2 は他の手法よりも高 SDR で分離する傾向が見られた．2DFilter はどの信号に対しても SDR は 0–2dB 程度であり，このパラメータセットではあまり分離に成功しなかった．

4.2　パラメータセットに関する考察

本実験により決定されたパラメータの傾向として，全体の SDR を最適にするパラメータセット (A) と $h(t)$ の SDR を最適にする (B) は概ね同様の傾向を示していることが観察できる．これは，通常の楽曲中においては打楽器的成分よりも調波的成分の方がパワーが大きいため，SDR の観点で見たときに，全体を最適化する場合も調波的成分にバイアスがかかるためと考えられる．一方，$p(t)$ の SDR を最適にする (C) では，$\kappa_1, \kappa_2, (\sigma_{\mathrm{H}}, \sigma_{\mathrm{P}})$ がこれらとは反対の傾向を示している．

反復回数に関しては，必ずしも多いほど性能が高くなるとは限らず，10 回程度の反復で計算を打ち切るほうが，収束するまで更新し続けるよりも有意に性能が高い場合が多く見られた．これは，目的関数を最適化することと SDR を最大化することは別の問題であり，過度に目的関数に小さくすることにより，$\boldsymbol{H}, \boldsymbol{P}$ が本来の音楽信号以上に滑らかなスペクトログラムになってしまい，信号に歪みが生じている可能性があると推測できる．

また，評価規準として SDR を用いていることに起因すると考えられる別の問題点として，(B) によって得られた $h(t)$ 成分，および (C) によって得られた $p(t)$ 成分が，聴感上は必ずしもそれぞれ調波音的のみ，もしくは打楽器的のみのようには聴こえない傾向にあることが挙げられる．すなわち，前者には打楽器的な成分が，後者には調波的な成分が聴感上は比較的多く残った．ただし，(B) によって得られた $p(t)$ および (C) によって得られた $h(t)$ では，調波音的な成分，打楽器音成分が抑圧され，特に HM2, Idiv ではほとんど聴こえなくなる傾向が見られた．また 2DFilter に関しても，SDR の観点からは必ずしも性能がよくないものの，聴感上は打楽器音がほとんど聴こえなくなるようなパラメータセットが存在することを確認している．したがって例えば和音認識などへの応用を考えたとき，(B) による $h(t)$ ではなく，(C) によ

Table 2　実験により決定した各手法のパラメータセット

手法	全体の SDR に基づくパラメータ (A)	調波的成分の歪みの小ささを優先したパラメータ (B)	打楽器的成分の歪みの小ささを優先したパラメータ (C)
HM1	$(\gamma = 0.4)$, $(\kappa_1 = 1.5)$, 反復 10 回, Wiener, $(L = 512)$	$(\gamma = 0.4)$, $\kappa_1 = 1.5$, 反復 10 回, Wiener, $(L = 512)$	$\gamma = 0.5$, $\kappa_1 = 0.25$, (反復 300 回), None, $L = 1024$
HM2	$\kappa_2 = 0.92$, (反復 10 回), None, $L = 1024$	$\kappa_2 = 0.92$, 反復 10 回, None, $(L = 512)$	$\kappa_2 = 1.04$, 反復 10 回, None, $L = 1024$
Idiv	$\sigma_{\mathrm{H}} = 0.5$, $\sigma_{\mathrm{P}} = 0.1$, (反復 300 回), Wiener, $(L = 512)$	$\sigma_{\mathrm{H}} = 0.5$, $\sigma_{\mathrm{P}} = 0.1$, (反復 10 回), Wiener, $(L = 512)$	$\sigma_{\mathrm{H}} = 0.1$, $\sigma_{\mathrm{P}} = 0.5$, (反復 300 回), Wiener, $L = 1024$
2DFilter	$(\gamma = 0.5)$, $c = 0.25$, $(I = 50)$, $(J = 10)$, Wiener, $L = 256$	$\gamma = 0.4$, $c = 1.5$, $I = 50$, $J = 50$, Wiener, $L = 256$	$\gamma = 0.5$, $(c = 1.2)$, $I = 50$, $J = 10$, Wiener, $L = 256$

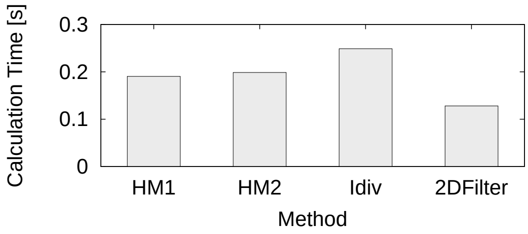

Fig. 3　各手法の計算時間の比較

る $h(t)$ を利用した方が，やや歪みはあるものの，打楽器音による妨害が小さいという利点があるため，和音認識精度の点では性能が高くなる可能性があると考えられる．

4.3　各手法の計算時間

これらの手法を実用する際には，分離性能のみならず計算時間も重要であることから，各手法に関する計算時間の評価を行った．いずれの手法も C++により同様の方法でプログラムを作成した．また、高速化のための工夫などは行わず，本稿にて述べた更新式をほぼそのまま実装した．実験に用いた計算機の CPU は Intel (R) Core 2 Duo (TM) P9400 2.40GHz である．HM1, HM2, Idiv の各手法の反復回数は 10 回に統一した．実験に用いたデータはサンプリング周波数 16kHz の長さ 10 秒のモノラル信号である．

Fig. 3 に各手法の計算時間を示す．なお、これらの時間は本稿で示した更新式等のみに要した時間を示し，ファイル入出力や短時間フーリエ変換の時間は含まれていない．いずれの手法も同程度のオーダーの計算時間であった．10 秒のデータを処理するのに要する時間は 10 秒より十分短く，実時間処理などへの応用を考える場合も，いずれの手法も十分に高速と考えられる．

5　まとめ・今後の課題

本稿では，著者らがこれまでに提案した調波音・打楽器音分離 (HPSS) の 4 手法を整理した．また，分離性能を SDR で評価し，これを最適にするようなパラメータの組み合わせを実験により決定した．2 次元フィルタによる時間周波数マスク設計に基づく手法 (HPSS-2DFilter) を除く 3 手法に関しては，概ね良好な分離性能を示すパラメータが決定できた．また，決定したパラメータに基づき，各手法の性能比較を行った結果，4 手法の中でも特に聴覚特性に近い滑らかさ規準と，分離信号が原信号に厳密に一致するという拘束条件に基づく手法 (HPSS-HM2) が，他の手法と比較してやや高い分離性能を示した．

HPSS では，通常は「調波的」と考えられるビブラート音や歌声などを「打楽器」に分離するようなパラメータセットがあることが分かっており，著者らはこれまでに HPSS を複数回適用することにより，これらの信号を強調する多重 HPSS という手法を提案している [5, 11]．多重 HPSS では HPSS よりもより多くの選択肢があるが，これらのパラメータの検討は今後の課題である．また，音楽鑑賞への応用には聴感上の音質が重要となるため，そのためのパラメータの設計も今後の課題となる．

謝辞　本研究の一部は日本学術振興会科研費特別研究員奨励費 (22-6961) の助成を受けて行われた．

参考文献

[1] Rigaud *et al.*, “Drum Extraction from Polyphonic Music Based on a Spectro-temporal Model of Percussive Sounds,” *Proc. ICASSP*, pp. 381–384, 2011.
[2] Ono *et al.*, “Harmonic and Percussive Sound Separation and its Application to MIR-related Tasks,” Springer 274, pp.213-236, 2010.
[3] Reed *et al.*, “Minimum Classification Error Training to Improve Isolated Chord Recognition,” *Proc. ISMIR*, pp.609-614, 2009
[4] Tsunoo *et al.*, “Rhythm Map: Extraction of Unit Rhythmic Patterns and Analysis of Rhythmic Structure from Music Acoustic Signals,” *Proc. ICASSP*, pp.185-188, 2009.
[5] Tachibana *et al.*, “Melody Line Estimation in Homophonic Music Audio Signals Based on Temporal-Variability of Melodic Source,” *Proc. ICASSP*, pp.425-428, 2010.
[6] Hsu *et al.*, “A Trend Estimation Algorithm for Singing Pitch Detection in Musical Recordings,” *Proc. ICASSP*, pp.393–396, 2011.
[7] 宮本 他, “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,” 音講論（秋）, pp.825-826, Sep., 2007.
[8] 宮本 他, “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, ”音講論（春）, pp.903-904, Mar., 2008.
[9] Ono *et al.*, “Separation of a Monaural Audio Signal into Harmonic/Percussive Components by Complementary Diffusion on Spectrogram,” *Proc. EUSIPCO*, 2008.
[10] Ono *et al.*, “A Real-time Equalizer of Harmonic and Percussive Components in Music Signals, ” *Proc. ISMIR*, pp.139-144, Sep., 2008.
[11] 橘 他, “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” 情処研報 MUS-81, No.12, 2009.
[12] 田口, “第 3 版実験計画法 上,” 丸善, 1995
[13] Vinyes, “MTG MASS database,” `http://www.mtg.upf.edu/static/mass/resources`, 2008.