多重 HPSS 法による混合音中の音声強調*

◎橘秀幸, 小野順貴, 嵯峨山茂樹 (東大院・情報理工)

1 はじめに

音声信号には，実環境においては様々な雑音が加わる．このような雑音は，音声認識などの音声アプリケーションの性能を低下させる原因となる．このため，これまでに様々な音声強調手法が提案されている [1, 2, 3, 4].

本稿では，スペクトログラムの時間周波数分解能を変えたときの音声と雑音の形状の変化に着眼し，本研究室で開発された調波打楽器音分離 (Harmonic/Percussive Sound Separation: HPSS)[5, 6, 7, 8] を多段階で用いる多重 HPSS 法により検出することにより音声を強調する，新しい手法を提案する．

2 多重 HPSS 法による音声強調

2.1 調波打楽器音分離 (HPSS) 手法

HPSS は，入力信号をスペクトログラム上で時間方向への連結が強い成分 (H) と，周波数方向への連結が強い成分 (P) とに分離する手法である．以下の 3 つの指針

- H と P の和は原信号に可能な限り一致
- H の時間方向の変化が可能な限り小
- P の周波数方向の変化が可能な限り小

に基づいて目的関数を設計し，その最適化により H,P を求める問題として定式化されている．なお，HPSS ではスペクトログラムの計算方法に関しては規定していないため，短時間フーリエ変換 (STFT) の窓関数やフレーム長，フレームシフトは任意である．

ある入力信号を HPSS により分離したときの H と P のエネルギーの比率を，H/P 比 (H/P Ratio) と呼ぶ．H/P 比は，ピアノやギターの音に関しては大きく，打楽器の音に関しては小さいことがこれまでの研究により分かっている．

2.2 様々な音響信号の H/P 比の傾向

音声信号には実環境においてはさまざまな種類の雑音が加わるが，特に理想化した雑音として，純音 (正弦波)，インパルス，ホワイトノイズの 3 種類に関する H/P 比について定性的に述べる．

正弦波はどのようなフレーム長で STFT を行った場合も，ある特定の周波数に高いピークが現れ，それが長時間保持される．すなわち，純音は時間方向への連結が強く，HPSS によりほとんどの成分が H に分離される ($H/P \gg 1$).

インパルスは，どのようなフレーム長で STFT を行った場合も，ある瞬間のみに全帯域を占め，その直前と直後の時刻には全くエネルギーがない．すなわち，インパルスは周波数方向への連結が強く，HPSS によりほとんどの成分が P に分離される ($H/P \ll 1$).

Table 1 各種信号に対する長短二種のフレーム長の HPSS の H/P 比の定性的な傾向.

	short frame	long frame
pure tone	very high (H)	very high (H)
impulse	very low (P)	very low (P)
white noise	medium	medium
speech	high (H)	low (P)

ホワイトノイズは，どのようなフレーム長で STFT を行った場合も，すべての時刻・周波数でほぼ等しいパワーを持つ．このため，時間方向と周波数方向の連結の強さは同等であり，HPSS を適用すると，H,P のどちらかへ極端にエネルギーが集中することはない．すなわち，H/P は中程度の値を示す ($H/P \approx 1$).

このように，正弦波，インパルス，ホワイトノイズは，いずれもフレーム長に関わらず H/P 比は同様の傾向を示す．

2.3 音声の H/P 比

音声には多くの場合に振幅やピッチにゆらぎ (短時間変動) が含まれるため，STFT のフレーム長に依存して異なった H/P 比の傾向が見られると考えられる．

音声を短いフレーム長 (10[ms] 程度) で STFT すると，フレーム内での音声は定常に近く，かつ周波数分解能が低いため，特定の周波数 bin にエネルギーが集中するため，周波数方向への連結性は弱く，HPSS により H へやや多く分離される ($H/P > 1$).

一方，長いフレーム長 (100[ms] 程度) での STFT の場合，フレーム内に音声の非定常な現象が含まれ，同時に周波数分解能が高くなるため，音声のピッチやパワーの変動に由来するスペクトルの広がりが複数の周波数 bin に亘って観測される．これにより，音声スペクトルでは周波数方向への連結が強く見られ，HPSS により P へやや多く分離される ($H/P < 1$)．このように，音声は STFT のフレーム長によって H/P 比の傾向が異なる (Table 1).

2.4 多重 HPSS 法による混合音中の音声強調

雑音と音声の H/P 比のこのような傾向の違いを利用すると，HPSS を 2 段階で用いることによって，音声を強調した信号が得られる．

すなわち，第 1 段階として，長いフレーム長のスペクトログラム上における HPSS により P 成分を抽出し，第 2 段階では，ここで得られた P 成分を一度波形に戻し，再び短いフレーム長のスペクトログラム上での HPSS によって H 成分を抽出する．

以上の 2 段階の処理において，音声はいずれの段

*Speech Enhancement in Mixed Audio Signals by Multi-stage Harmonic-Percussive Sound Separation (HPSS), by TACHIBANA Hideyuki, ONO Nobutaka and SAGAYAMA Shigeki (Graduate School of Information Science and Technology, The University of Tokyo).

階も通過しやすいのに対し，正弦波，インパルス，ホワイトノイズに類する雑音はいずれかの段階において抑圧される．これにより，音声強調信号が得られると考えられる．

この分離手法を多重 HPSS 法と呼ぶこととし，この手法を本稿で提案する音声強調手法とする．

3　音声強調の性能評価

3.1　実験目的と条件

提案手法の性能を評価するため，音声信号に雑音を加算した信号に対し多重 HPSS 法を適用し，SNR の改善値を調べた．

音声信号には，SMILE2004 データベース [9] より，男女それぞれの日本語の朗読音声を使用した．雑音には，同データベースから，ホワイトノイズ，ピンクノイズ，1 kHz の帯域雑音，足音（ボイド 280 素面，ハイヒール強め歩行），ドライヤー，地下鉄車内騒音，弦楽四重奏，赤ん坊の泣き声，合唱，池袋駅地下コンコースの 10 種類と，計算機上で生成した 1kHz の正弦波，1[s] 間隔のインパルス，の計 12 種類を用いた．音声と雑音は，10 [s] を切り出して使い，-10 dB から 10 dB まで，2.5dB 刻みの SNR で混合した．

多重 HPSS 法に用いたフレーム長は，第 1 段階 HPSS では 256[ms] とし，第 2 段階 HPSS では 16[ms] とした．それぞれの HPSS において，STFT に用いた窓関数は，分析窓，合成窓ともにハニング窓の平方根とした．

3.2　音声強調の実験結果

女性の朗読音声に各雑音を加算して提案手法を適用したときの SNR の改善値を Fig. 1 に示す．なお，男性の朗読音声に関しては，ここで示した値より 1 dB 程度ずつ低い値を示した．

今回用いた 12 種類の雑音の中では，正弦波とインパルスに対して高い改善値を示した．次いで足音，ホワイトノイズ，ピンクノイズ，地下鉄車内騒音，ドライヤーの順に高い改善値を示した．ほとんどの雑音に関してある程度の効果が認められたが，合唱や赤ん坊の泣き声のように雑音自体も音声である場合は，SNR の向上は認められなかった．

4　まとめ

本稿では，音声の調波構造や入力信号の SNR などの事前知識を用いることなく，スペクトログラムの時間周波数分解能を変えた 2 段階の分析のみによって音声を強調する手法を提案した．また，雑音と音声の混合信号を用いた評価実験により，複数の種類の雑音に対する効果を確認した．

今後は，HPSS の階層を直列にさらに増やすことや HPSS の並列化による本手法の改良の検討，本手法を音声認識などへの前処理とした場合の性能評価などを進めていく予定である．また，同様の手法で音楽中から歌声を抽出できることが分かっている [10] が，本稿の手法との統合や一般化についても，今後の検討課題となる．

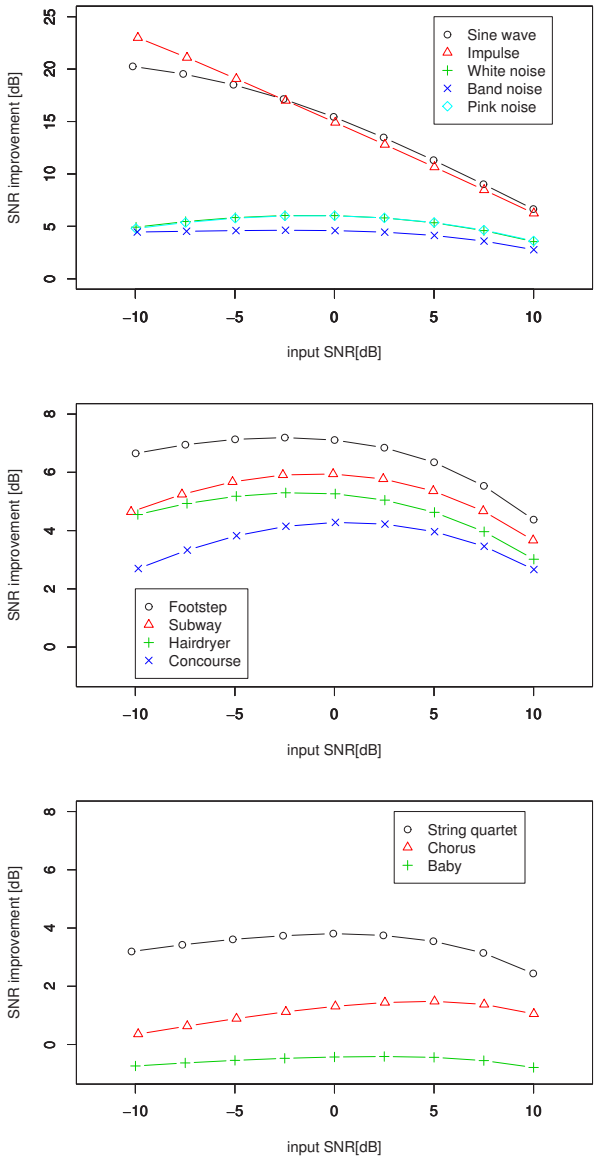

Fig. 1　入力信号の SNR と，多重 HPSS 法適用による SNR の改善値（音声はすべて女声）．

参考文献

[1] Berouti *et al.*, *Proc. of ICASSP,* pp.208-211, 1979.
[2] Boll, *IEEE Trans. ASSP,* Vol.27, No.2, pp.113-120, 1979.
[3] Ephraim, Malah, *IEEE Trans. ASSP,* Vol.32, No.6, pp. 1109-1121, 1984.
[4] Cohen, Berdugo, *Signal Processing,* Vol. 81, pp. 2403-2418, Elsevier, 2001.
[5] Ono *et al.*, *Proc. of ISMIR*, pp.139-144, 2008
[6] Ono *et al.*, *Proc. of EUSIPCO*, 2008.
[7] 宮本他, 音講論 (春), pp.903-904, 2008.
[8] 宮本他, 音講論 (秋), pp.825-826, 2007.
[9] DVD 版 建築と環境のサウンドライブラリ，日本建築学会編，技報堂出版 (SMILE2004)
[10] 橘他, 音講論 (春), 2-8-8, 2009.