

ドラム演奏ミキシングのための漏洩音抑制

村松 知哉

Tomoya MURAMATSU

法政大学情報科学部デジタルメディア学科

E-mail: tomoya.muramatsu.7b@stu.hosei.ac.jp

Abstract

The purpose of this research is to reduce Microphone-leakage, which is an important problem in drum recording. Mixing a drum sound source with a lot of Microphone-leakage requires expertise and experience, and it is difficult to mix each instrument clearly in the case of a sound source with a lot of leakage sound. Therefore, in this research, we propose a method for separating the Microphone-leakage of microphones with fast-attack percussion instruments and low distortion, which is specialized for drum recording and mixing, by extending the multi-channel Wiener filter. It improves the separation performance when the leakage signal is louder than the acquisition signal by taking the correlation between the acquisition signals and using the values as weights. In addition, by varying the STFT window length, we reduced the pre-echo that often occurs in the transient section and improved the SDR. The results of separation and mixing with various drum sources showed that the proposed method effectively suppressed the leakage sound of percussion instruments.

1 まえがき

ドラムは楽曲においてリズムセクションの役割を担っており、バンドアンサンブルの土台となる重要な楽器といえる。他の楽器と異なるドラムの主な特徴として、1. 音階がない 2. アタック音がとても強い 3. 複数の楽器の集合体で独りで演奏する、の3つが上げられ、ドラム専用の録音方法や音声波形編集法が存在する。

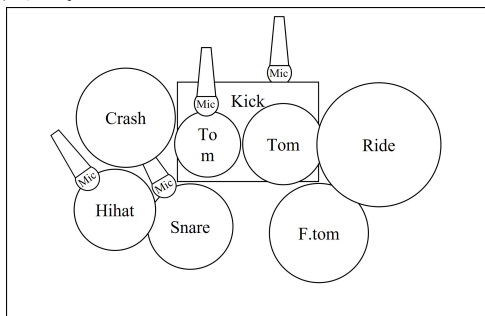


図 1. スネアマイクに多く入る漏洩音を分離するためのマイキング方法

ドラムは一般に、図 1 に表すように大太鼓・中太鼓・小太鼓・シンバルなどの一式からなる。また図 1 はスネアマイクに多く入る漏洩音を分離するためのマイキング例である。通常のドラム録音のマイキング方法は、各楽器に最低一本のマイクを設置、そしてドラムセット全体を録音するマイクをセット上部に設置

する。またエンジニアによっては追加で録音空間の残響音を収音するアンビエントマイクを複数設置する場合もある。

このドラム録音で問題となる現象が Microphone Leakage(マイクの漏洩音)[6] である。この漏洩音とは同じ空間に複数の楽器とマイクが存在するとき、意図していない楽器音がマイクに収音されてしまう現象のことである。ギターやピアノなどの楽器はそれ単体で楽器として成立し、録音時にはそれぞれ別の録音ブースに分けたり、楽器パート毎に録音することでこの漏洩音は解消される。しかしドラムは複数楽器の集合体であるため、先に述べた方法が行えず漏洩音を避けることが出来ない。図 2 は一般的なスネアマイクの短時間 RMS である。スネアドラムの音量は -12dB を推移し、バスドラムやシンバル類の漏洩音は -25dB を推移している。このマイク信号でスネアの音色を処理しようとしても漏洩音の影響が大きいため、スネア単体の処理には専門的な知識や経験が必要となる。そして初級者の演奏の録音では漏洩音が特に大きく、ミキシングで個々の楽器を処理することが難しくなってしまう。

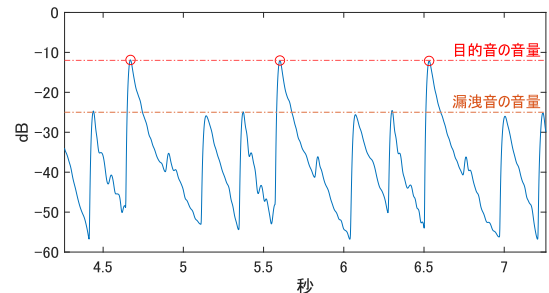


図 2. スネアマイクの短時間 RMS

2 漏洩音の分離手法と問題点

ドラム録音をはじめとした楽器演奏音を分離する方法にブラインド信号源分離 (BSS) やウィーナフィルタ、ニューラルネットワークを活用した方法が存在し、このとき問題となるのが、統計独立性と反響音-リバーブの影響である。演奏をする場合、同じ音程やハーモニーが同時に鳴ることが多く、BSS で必要となる統計独立性が保てなくなる。また BSS はリバーブの影響で性能が悪くなり、ドラム録音のような空間の反響も録音対象となる場合には向かない。ニューラルネットワークを活用した方法ではリアルタイム処理用に考えられているものが少なくミキシング、ライブ PA における使用は難しい。また先行研究 [2] では信号対干渉比 (SIR) の性能を重視して評価しており、位相歪みをはじめとした歪みが多くミキシング作業には向かないという現状があった。

ウィーナフィルタは、漏洩音が目的音源に比べ小さく無視できることを前提としている [1]。しかしシンバルの音を録音するためのマイク (ハイハットマイクやオーバーヘッドマイク等) では、太鼓類の音量が大きいため、その前提条件を達成できない。そのため一部のマイクの処理性能が悪くなる。

そして STFT で窓を長くし周波数分解能を上げると、打楽器の波形では”Pre-echo(図 3 点線部分)”や”アタックの滲み”と呼ばれる現象が起こる。STFT では窓内の信号が定常であることを前提としており、トランジェント区間の急峻な波形の変化ではその定常性が崩れてしまうことにこの現象は起因する。Pre-echo は楽器音が発生する直前のノイズとして現れ、これによってアタックがぼやけ、音のシャープさが失われるように感じてしまう。

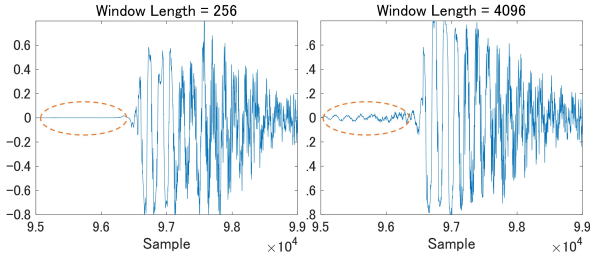


図 3. 窓長を変えて処理した音源

本研究では、先行研究で使用され反響音に左右されづらく統計独立性が必要でないウィーナフィルタを拡張、また新たにトランジェント検出置換、漏洩音が無視できない収音信号の処理を考案する。そしてミキシングの手助けとなるドラム録音に特化したマイクの漏洩音分離の手法を新たに提案する。

3 提案手法

今回提案する手法の処理の流れ(図 4)は、まず入力された信号にマイク同士の遅延の補正を行い、ABS-STFT(Adaptive Block Size-STFT)[7]で時間周波数領域に変換する。そしてウィーナフィルタを推定し信号へのフィルタ適用を行う。ウィーナフィルタのための時間領域の前処理として、瞬時チャンネル間相関推定、ドミナント周波数-PSD 推定を用いる。最後にトランジェント検出を行い、検出されたフレームをウィーナフィルタに通過させないことによってアタック音が歪むことを防ぐ。

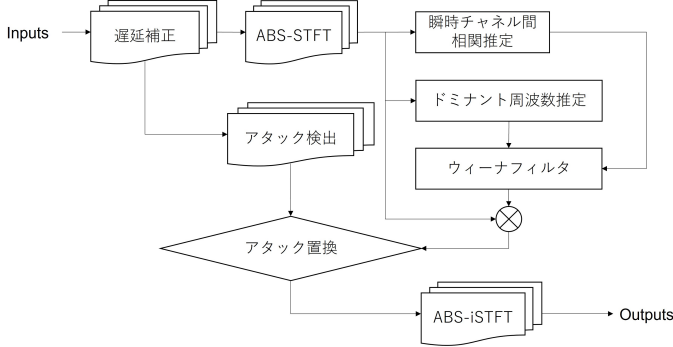


図 4. 提案手法のフローチャート

3.1 マイク間の遅延補正

マイク間の距離による遅延はミキシングやウィーナフィルタの推定において誤差を生じさせる原因の一つである。この遅延を補正するため、収音信号同士の相互相関を計算し一番楽器から離れているマイクの遅延に他のマイクを合わせるディレイ機能を作成する。

3.2 アタック検出と置換

ドラムのアタックに何かしらの音声処理を行うとアタックの歪みが起こる。これを回避するためにアタック検出 [4] を行い、アタックのみフィルタ処理を通過させない機能を作成。

3.2.1 アタック検出

アタックとは複数の周波数帯域で顕著なエネルギーピークが出現する短い領域の事である。図 5 はスネアドラムの収音信号

の短時間 RMS であり、 -10dB 付近を推移する赤丸部分がスネアドラムである。収音信号にハイパスフィルタを掛け、各周波数帯域で複数の閾値を超える周波数帯域が検出された場合、その箇所をアタックポイントとする。またこのときの閾値の値は収音信号によって異なるが、図 5 では閾値を -15dB として検出を行っている。

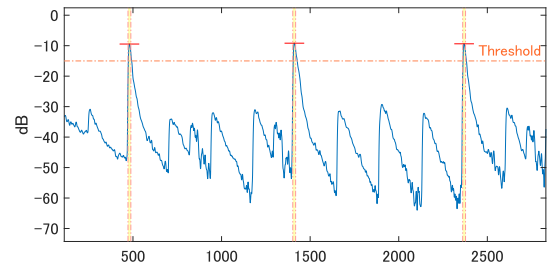


図 5. アタックポイントと閾値 (-15dB)

3.2.2 アタック置換

上記の方法で求めたアタックポイントをウィーナフィルタ処理前の収音信号に置き換える。このときスネアの音色やディケイの長さによって、アタック置換を行う区間の長さの変更を行う。

3.3 ABS-STFT

音の立ち上がり早い楽器を STFT する場合、窓サイズを小さくすることで時間分解能が向上し音の立ち上がりを認識できるが、その代わり周波数分解能が低下する。そのため一つの周波数ビンで複数の倍音を表さなければならず、ミキシングを行うには音源の解像度が荒くなる。逆に窓サイズを大きくすると周波数分解能は向上するが時間分解能が低下し、前述した Pre-echo が発生する。

今回は Pre-echo の影響を少なくして周波数分解能を向上するために ABS-STFT(Adaptive Size Block-STFT)を使用する。これは Pre-echo が発生するトランジェントポイント(図 6 の赤い四角形部分)を検出し、その区間のみ STFT の窓長を短くするというものである。

このトランジェントを検出するために対象信号の短時間 RMS の変化量に着目し、閾値を超える地点をトランジェントポイントとする。

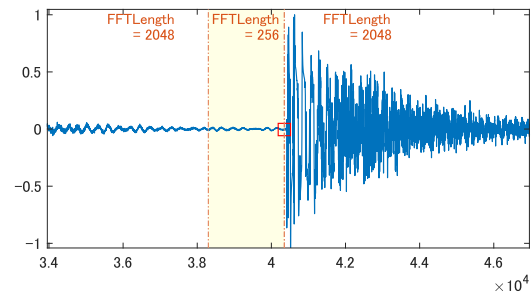


図 6. トランジェント検出ポイント(四角形部分)と窓長を変更する区間

3.4 マルチチャンネルウィーナフィルタ

本研究のウィーナフィルタは先行研究で導出されたマルチチャンネルウィーナフィルタ(式 1)を使用する。これは通常のシングルチャンネルウィーナフィルタを拡張する形 [1] で作成されるもので、フィルタ単体ではマイクの本数を増やしていくほどフィルタ性能が著しく落ちてしまう。そのため”ドミナント周波数-PSD 推定”で述べる拡張手法を提案し、フィルタ性能を向上させる。

$$W(\kappa, \omega_k) = \frac{P_{s_m, s_m}(\kappa, \omega_k)}{P_{s_m, s_m}(\kappa, \omega_k) + \sum_{i=1}^M P_{u_i u_i}(\kappa, \omega_k)} \quad (1)$$

ウィーナフィルタを用いた分離式は以下の通り。

$$Y_m(\kappa, \omega_k) = W_m(\kappa, \omega_k) X_m(\kappa, \omega_k) \quad (2)$$

また第2節で述べたように、オーバーヘッドマイクなどの一部マイクでは漏洩音が目的音と同等の音量で無視できない。そのため先行研究ウィーナフィルタの前提条件を満たせなくなり、どれだけ漏洩音が分離できているかを表す SIR(信号対干渉比)が悪くなる。これを回避するために瞬時チャンネル間相関推定という各マイク收音信号の重み付けを行う。

3.5 ドミナント周波数-PSD 推定

ウィーナフィルタ通過前後で変化のない周波数領域は漏洩音の影響を受けていない純粋な收音信号といえる。そしてその周波数領域に重みを付けたパワースペクトル密度を推定する。そのパワースペクトル密度を重みづけ関数(3)としてフィルタに適用することで、ウィーナフィルタの処理精度を向上させる。

$$P_{s_m, s_m}(\kappa, \omega) = \gamma_m(\kappa) P_{s_m, s_m}(\kappa - 1, \omega) + (1 - \gamma_m(\kappa)) [P_m^D(\kappa, \omega) + P_m^R(\kappa, \omega)] \quad (3)$$

3.6 瞬時チャンネル間相関推定

ある区間において、”目的音を奏でる楽器の收音信号” s 内の漏洩音が大きくなり無視できなくなるという状況は、結果的に”漏洩音となっている楽器の收音信号” u に s が相似しているということとなる。つまり漏洩音の影響が大きい区間の相関係数は大きくなり、影響が小さければ無相関になるということである。図7はハイハットマイクとスネアマイクの瞬時相関を求めたグラフで、スネアの発音区間で相関が大きくなっていることが分かる。そして收音信号をオクターブフィルターバンクで分け、各周波数バンド ω 毎に收音信号間で相関を求め、瞬時チャンネル間相関行列 H [5] を作成する。その行列を重みづけ関数としてフィルタに適用する。

$$s_{\kappa\omega} = \begin{bmatrix} s_{\kappa_1\omega_1} & \cdots & s_{\kappa_M\omega_1} \\ \vdots & \ddots & \vdots \\ s_{\kappa_1\omega_N} & \cdots & s_{\kappa_M\omega_N} \end{bmatrix}$$

$$H = s_{\kappa\omega} u_{\kappa\omega}^H = \begin{bmatrix} s_{\kappa_1\omega_1} u_{\kappa_1\omega_1}^* & \cdots & s_{\kappa_M\omega_1} u_{\kappa_M\omega_1}^* \\ \vdots & \ddots & \vdots \\ s_{\kappa_1\omega_N} u_{\kappa_1\omega_N}^* & \cdots & s_{\kappa_M\omega_N} u_{\kappa_M\omega_N}^* \end{bmatrix}$$

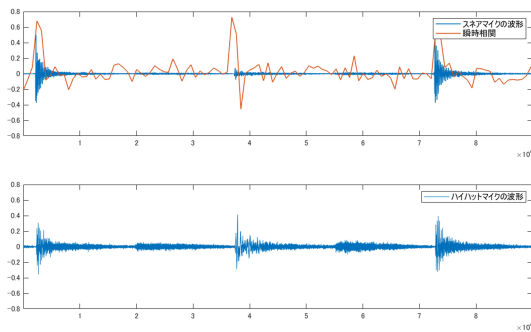


図7. スネアとハイハットマイクの瞬時チャンネル間相関

4 評価方法

本研究はどれだけ漏洩音が分離できているかを表す SIR(信号対干渉比)、どれだけ信号が歪んでいるかを表す SDR(信号対歪み比)、そして主観評価実験の4つの評価方法を使用する。

$$y_n = s_{\text{target}} + e_{\text{interf}} + e_{\text{noise}} + e_{\text{artif}}$$

$$\text{SIR} = 10 \log_{10} \frac{\|s_{\text{target}}\|^2}{\|e_{\text{interf}}\|^2}$$

$$\text{SDR} = 10 \log_{10} \frac{\|s_{\text{target}}\|^2}{\|e_{\text{interf}} + e_{\text{noise}} + e_{\text{artif}}\|^2}$$

$$\text{SAR} = 10 \log_{10} \frac{\|s_{\text{target}} + e_{\text{interf}} + e_{\text{noise}}\|^2}{\|e_{\text{artif}}\|^2}$$

実験方法としては、ドラムフレーズを実際に叩きマルチトラックで録音し、その後各楽器単体音をそれぞれ收音する。ドラムフレーズと各楽器単体音源を同じフィルタにかけ漏洩音を分離し、分離後ドラムフレーズは主観評価実験に使用する。分離後楽器単体音源は評価計算に使用し、SIR、SDR、分離前後の SIR Improvement[3] を計算。そして計算結果を提案手法と先行研究手法で比較する。今回は3種類の音色と2種類の音量バランス、2つの分離難易度の計12種類(表1)のドラム演奏を收音した。

表1. ドラムの演奏パターン

音色	音量バランス	サステインの長さ
ハイピッチ	シンバルの音量大	長い
ミドルピッチ	太鼓類の音量大	短い
ローピッチ		

主観評価実験では、先行研究手法と提案手法で分離された音源を同じパラメータでミキシングする。そして分離を行っていないミキシング音源を参照音源とし、2つの分離音源のどちらがより歪みのない音源かを評価する。[8][9] またこの音源はマルチトラックドラム音源*1の中から3種類の長さのサステインの音源(2種類は上記の長いサステインに、1種類は短いサステインに分類される)を使用した。

5 結果

5.1 性能評価

それぞれ評価したものが(図8上)となっている。全ての場合で20dB以上 SIR が向上しており、先行研究手法(図8下)に比べて5dB以上の向上がみられる。SDRについては先行研究手法とあまり変わらずどちらも8dBを推移する結果となった。

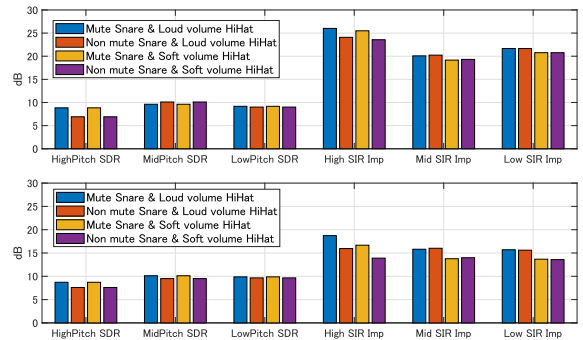


図8. 提案手法(上)と先行研究(下)の性能評価比較

*1 Drumdrops, "Drum Loops Tracks", <https://www.drumdrops.com/>

今回提案した補正項目の性能を評価する。アタック検出置換、ABS では SIR、SDR どちらも変化は少なかった。瞬時チャンネル間相関は SIR が 10.4dB と大幅な向上、SDR は -4.4 dB の低下となった。また、SDR の変化については歪み自体が信号にマスキングされている場合などが存在するため、数値だけで判断することは難しい。また処理後のサステインの長さが短くなるなど、音質に変化があると聴覚上感じたため、主観評価実験を行って音質の変化を比較した。

5.2 主観評価実験

主観評価ではミキシングや作曲編曲、音楽活動の経験がある 20 人を被験者とした。結果はサステインの長いドラムでは 76%、短いドラムでは 88% の被験者が提案手法を歪みのない音源と判断した。(図 9)。また、サステインの長いドラム音源では参照音源と提案手法で音色が異なるという感想が上がった。

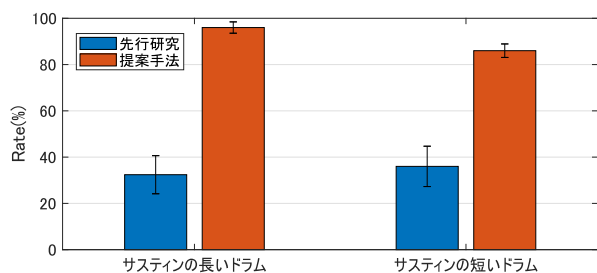


図 9. 主観評価実験結果

6 考察

SIR の約 20dB の向上はミキシングのコンプレッション処理の平均的なリダクションレベルである -6 dB を大幅に上回っており、過激なコンプレッション処理など積極的な音作りが十分可能な分離性能が得られた。

それぞれの補正項目の評価においてはアタック検出置換と ABS の計算上の性能が向上しなかったが、SDR の変化については歪み自体が信号にマスキングされている場合などが存在するため、数値だけで判断することは難しい。

また瞬時チャンネル間相関においては SIR の大幅な向上と SDR の低下が起こった。これは、目的音から離れているマイクと目的音のオンマイクで相関を求め重みづけを行ったことが原因と考えられる。太鼓類のオンマイクポジションは、演者がマイクを叩いてしまうことを防ぐために、打面の端に位置する 경우가多い。そのためマイクが太鼓類のリリースを收音しやすいポジションとなり、オフマイクの收音信号と音色の変化が起こる。その結果リリース区間のスペクトル構造が変化することで相関が小さくなり、これによってリリース区間において目的音もフィルタで遮断してしまう。

先行研究と提案手法において SDR の向上がなかった理由として、ドラムセット内の位相特性が極端に変化することが原因と考える。図 10 は従来行われていた漏洩音分離実験環境とドラムセット内でそれぞれ測定した位相応答である。従来実験環境では位相応答がリニアであるのに対して、ドラムセット内では崩れている。これによってパワースペクトルで計算を行うウィナーフィルタやチャンネル間の位相に左右される瞬時チャンネル間相関の動作が安定しなくなった。

処理速度においては音質を担保するために長い窓長を必要とし、新たな補正項目を増やした結果、処理が重くなることが問題点に上がった。

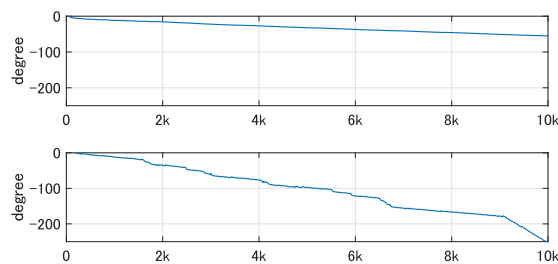


図 10. 従来評価実験のオフマイク位相応答 (上) とドラムセット内のオフマイク位相応答 (下)

7 あとがき

本研究では、チャンネル間の相関関係とトランジェント-アタック区間への新たなアプローチから打楽器の漏洩音分離手法を提案した。評価実験から今回の提案手法で SIR が約 20dB 向上し、ミキシングに十分な音質と分離性能で分離を行えることが分かった。この手法でミキシングの手間を減らすことが可能であると考えられる。しかし音質を担保するためには長い窓長を必要とし、新たな補正項目を増やした結果、処理が重くなることが問題点に上がった。システムにおいてはまだ改良する余地があると考えられ、今後はリアルタイムで処理が可能なアルゴリズムの見直しを行っていきたい。

参考文献

- [1] E. K. Kokkinis, J. D. Reiss and J. Mourjopoulos, “A Wiener Filter Approach to Microphone Leakage Reduction in Close-Microphone Applications,” in *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 3, pp. 767-779, March 2012.
- [2] E. K. Kokkinis and J. Mourjopoulos, “Unmixing acoustic sources in real reverberant environments for close-microphone applications,” *J. Audio Eng. Soc.*, vol. 58, no. 11, pp. 1710, Nov. 2010.
- [3] E. Vincent, R. Gribonval and C. Fevotte, “Performance measurement in blind audio source separation,” in *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 4, pp. 1462-1469, July 2006.
- [4] X. Rodet and F. Jaillet, “Detection and modeling of fast attack transients,” in *ICMC, 2001, NA, France*. pp.1-1.
- [5] H. Sawada, H. Kameoka, S. Araki and N. Ueda, “New formulations and efficient algorithms for multichannel NMF,” *2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2011.
- [6] E. Kokkinis, A. Tsilfidis, T. Kostis, and K. Karamitas, “A New DSP Tool for Drum Leakage Suppression,” *Engineering Brief 108*, (2013 October.).
- [7] Imen Samaali, et.al. “Attack restoration in low bit-rate audio coding, using an algebraic detector for attack localization,” *ISVC, Sep 2010, Rabat, Morocco*.
- [8] E. Vincent, M. Jafari, and M. Plumbley, “Preliminary Guidelines for Subjective Evaluation of Audio Source Separation Algorithms,” in *Proc. ICA Research Network International Workshop (2006 Sept.)*.
- [9] M. Lee, I. Heo, N. Choi, and K. M. Sung, “On Evaluation of Blind Audio Source Separation,” in *Proc. AES 34th Int. Conf. on New Trends in Audio for Mobile and Handheld Devices (Jeju Island, Korea, 2008 Aug.)*.