

ブロードバンド配信コンテンツにおける特定音楽情報除去手法の検討

松岡正悟[†] 高木真一[†] 小舘亮之[†] 富永英義^{†,††}

[†] 早稲田大学 国際情報通信研究センター
〒 169-0051 東京都新宿区西早稲田 1-3-10
^{††} 早稲田大学 理工学部
〒 169-8555 東京都新宿区大久保 3-4-1
E-mail: †matsu@tom.comm.waseda.ac.jp

あらまし マルチメディアコンテンツ配信において、そのコンテンツに含まれる様々な著作権の管理は困難である。本手法では、著作権などの理由により特定音楽情報をインターネット配信できないものを音源分離手法を用いて除去するコンテンツ配信システムを提案する。コンテンツの混合音に含まれる抽出したい音楽は、音源分離の際のリファレンス音源として保持していることを前提とし、入出力信号を用いて伝達関数の推定を行い音源分離を実現する。

キーワード 著作権保護, マルチメディアコンテンツ配信, 音源分離

Study of Audio Separation Method on Contents Distribution

Shogo MATSUOKA[†], Shin'ichi TAKAGI[†], Akihisa KODATE[†], and Hideyoshi TOMINAGA^{†,††}

[†] GITI, Waseda University
29-7 Building 1-3-10 Nishi-Waseda, Shinjuku-ku, Tokyo, 169-0051 JAPAN
^{††} Dept. of Science and Engineering, Waseda University
3-4-1 Ohkubo, Shinjuku-ku, Tokyo, 169-8555 JAPAN
E-mail: †matsu@tom.comm.waseda.ac.jp

Abstract With the increasing low-cost, high performance digital video camera, other hardware and broadband network, it has become easier to create and distribute multimedia contents. However, it is very difficult to attend content distribution using personal library(music, image, logo,etc) may violate content holder's copyright. As a solution for this problem, in this paper we propose a new approach for music copyright protection on content distribution. The basic idea of this approach is to use audio separation techniques using reference music source which is included in multimedia content. The algorithm is presented transfer function estimation from between actual output and reference music source's output signal.

Key words Digital Rights Management, Multimedia Content Distribution, Audio Source Separation

1. はじめに

近年のブロードバンドネットワークの普及，ストリーミング技術や圧縮技術の発展，またデジタルビデオカメラ，デジタルスチルカメラ，コンピュータ，そのオーサリングアプリケーションなど，ハードウェア・ソフトウェアの高性能化，低価格化によりデジタル音楽や画像，動画，本やゲーム等のマルチメディアコンテンツがインターネットを通じて末端のユーザに対して容易に提供できるようになった．これらの変化は，ユーザによる容易なマルチメディアコンテンツの制作・配信・蓄積の可能性を示唆している．その一方で，デジタルコンテンツはデータの劣化なくコピーを行うことが出来るため，著作権に絡んだコンテンツの無断使用などの問題も生じてきており，デジタルマルチメディアコンテンツの著作権管理 (Digital Rights Management) に注目が集まっている．

一般ユーザが制作したコンテンツ中に，そのユーザが配信権を持たないオブジェクトが存在する場合，そのコンテンツをインターネット配信することは違法行為となり，出来ない．そのため，このようなコンテンツのインターネット配信を実現するためには，著作権情報を含む映像や音楽をコンテンツ内から取り除く等の情報加工が必要となる(図1)．本検討では，特に音楽信号に注目し，マルチメディアコンテンツに含まれる配信権のない音楽信号を音源分離手法を用いて除去するコンテンツ配信システムを提案する．音源除去処理には，混合音に含まれる抽出したい音楽信号の出力元となった，CD や DVD 等のリファレンス信号を用いて，クロススペクトル法により伝達関数の推定を行い，配信権のない音楽信号だけをコンテンツ内から取り除く．

本稿では，2. において背景として提案システムの要素技術であるデジタルコンテンツ著作権管理システム，音源分離手法に関する先行研究，その現状の課題を示し，3. において提案方式を述べる．また，4. では提案システム内の音源分離プロセスに関する詳細な説明を行い，5. において提案手法を適用したシミュレーション・その評価を行った．最後に，6. においてまとめと今後の課題について述べる．また，音の定義は人の声などの音声を Speech，音楽を Music，雑音などの音を Sound と定義する．

2. 背景

2.1 権利の契約

マルチメディアコンテンツの大量流通により，コンテンツに含まれる音楽著作権，画像中の人物肖像権など，配信権がユーザにないオブジェクトの取り扱いが問題になることが予想される．これらのコンテンツに含まれる著作権情報を保護するための技術・研究として Digital Rights Management(DRM) 技術 [1] やコンテンツ ID フォーラム (Contents ID Forum) [2]，電子透かし [3] が挙げられる．オーディオの分野に注目すると，音楽に含まれる固有情報を Fingerprinting として検出することにより，コンテンツ同定を行いコンテンツ管理を行う研究も行われている [4] [5]．しかし，これらのアプローチでは一般ユーザによるコンテンツ配信を妨げることにもつながり，問題の根本的な解決にはまだ多くの課題が残されている．

2.2 Digital Rights Management 技術

従来，デジタルコンテンツをユーザ同士がやり取りする場合は，何度コピーしてもどの様な遠距離を送受信しても品質が劣化しないという，デジタルの特性により生じた著作権など権利問題から，DRM 技術等のデジタルコンテンツの流通・再生に制限を加える技術を用いていた．その具体的な実装形態は様々で，メモリカードなどの記憶媒体に内蔵される場合や，音

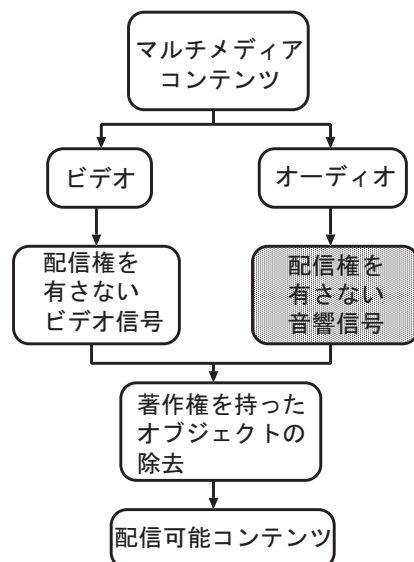


図1 ユーザの立場に立った著作権保護

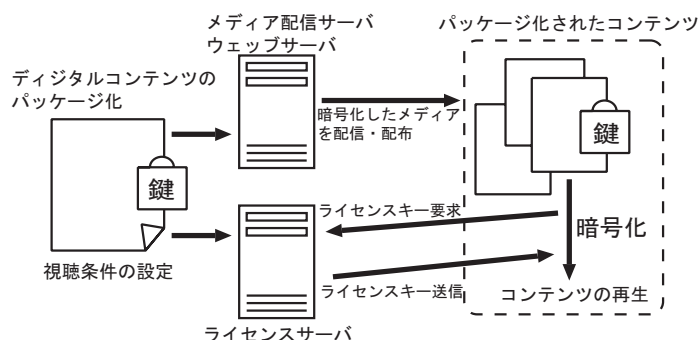


図2 Digital Rights Management システムの概要

声や動画のプレーヤーソフトに組み込まれる場合，送受信・転送ソフトに組み込まれる場合，およびそれらの組み合わせなどがある．

図2. は従来手法における情報配信システムを説明する図である．コンテンツオーナーはそのコンテンツに対してライセンスキーを付与し，暗号化されたファイルフォーマットで配布する．その情報の利用者がコンテンツを利用する際，ライセンスサーバにライセンスキーを要求すると，ライセンスサーバは利用者の ID を認証し，正規ユーザの場合にはライセンスを発行する．ユーザは，ライセンスキーにより情報を開封し，コンテンツオーナーが設定した条件の下で利用が可能になる．

しかし，これらのアプローチでは，容易なユーザによるマルチメディアコンテンツの配信が困難であるという面で，一般ユーザによるコンテンツ配信を妨げることにもつながり，問題の根本的な解決にはまだ多くの課題が残されている．

2.3 音源分離技術

複数の音が混在する中で，自分が望む信号だけを分離抽出する問題は音源分離と呼ばれ，工学的な処理で実現するには難しい問題であり現在でも盛んに研究が行われている．これは，混合信号から目的の信号を抽出するという問題において，個々の信号がどのように混合されたのかを表す情報が欠落していることに起因する．

現在，盛んに研究が行なわれている音源分離手法としてブラインド音源分離法 (Blind Source Separation Method) が挙げられる [6] [7] [8]．その音源分離プロセスを図3 に示す．ブラインド音源分離とは複数の音源信号が混在して観測される場合，観測

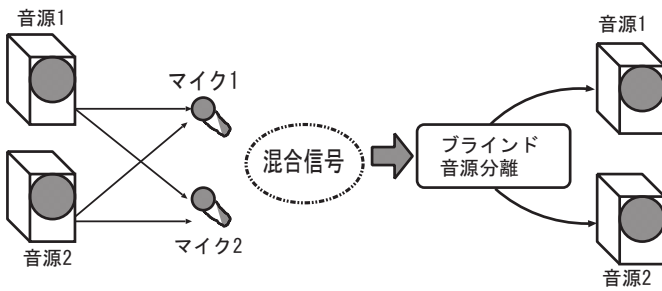


図3 ブラインド音源分離

信号のみから音源信号を推定する技術である。このブラインド音源分離に関する研究では、独立成分分析 (ICA) と呼ばれる音源の独立性を条件にした手法に基づいたものが主流になっている [6]。

従来研究されてきた ICA を用いたブラインド音源分離手法として、時間領域 ICA (TDICA) や周波数領域 ICA (FDICA) が挙げられる。TDICA は時間領域において FIR 型の音源分離フィルタを推測し、FDICA は周波数領域において各周波数毎に音源分離フィルタを推測する手法である。しかし、これらの手法では高残響実環境下では十分な性能が得られなかった。そのため高残響実環境にも対応する多段 ICA (MICA) も西川らによって検討されてきた [8]。

2.4 既存技術の欠点

MICA, TDICA, FDICA 共通の解決困難な問題として挙げられるのが、音源同士の相関性が高い、音源が多数存在するなどの混在条件が複雑な場合の音源分離である。そのため、様々な音場が存在すると考えられるマルチメディアコンテンツでは、配信権をユーザが持たない音楽情報を持つマルチメディアコンテンツ除去という目的には、ブラインド音源分離手法は適さないと考えられる。

3. 提案手法

本稿では配信したいコンテンツに含まれる、ユーザが配信権を持たない音楽を取り除くことを目的に、音源分離手法の検討を行なう。まず、マルチメディアコンテンツをビデオとオーディオに分離し、オーディオ情報だけを用いて音源分離処理を行なう。さらに、処理後のオーディオ情報に著作権情報をもたない音楽を付加し、再びビデオと合成する。

想定するシステムの流れを以下に示し、そのイメージを図4示す。

3.1 システムイメージ

- (1) Input
著作権を持った音楽を含むマルチメディアコンテンツの
入力
- (2) Demultiplexer (Demux)
ビデオストリームから映像と音に分離
分離後の音情報は wave フォーマットとして保持する
- (3) Matching
著作権に触る音楽とレファレンス音との同期
- (4) Audio Separation By Using Reference
レファレンス音をもとに混合音から目的の音楽
を抽出する
- (5) Insert Copyright-Free Music
削除した部分に著作権に触らない音楽を入れる
- (6) Multiplexer (Mux)
分離している音と映像を同期させまとめる
- (7) Output

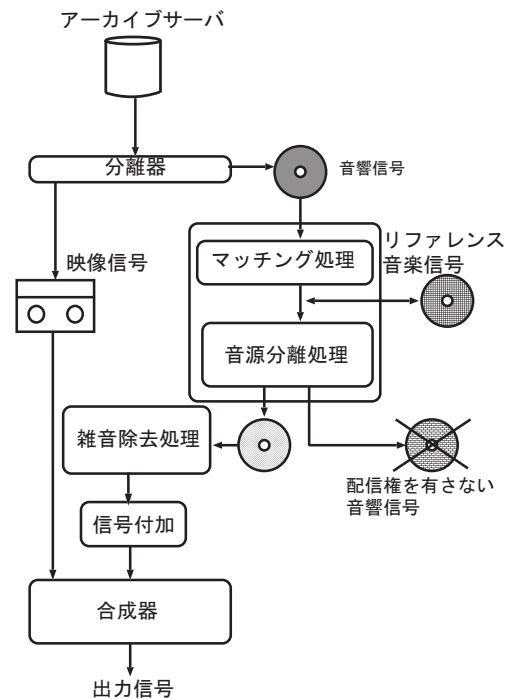


図4 コンテンツ配信サーバ内のシステムイメージ

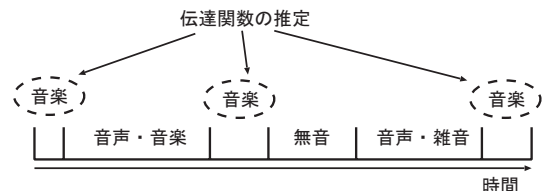


図5 マルチメディアコンテンツのオーディオ情報

システム出力のマルチメディアコンテンツの出力

3.2 想定するマルチメディアコンテンツ

マルチメディアコンテンツは映像情報と音情報を含んでおり、音情報に注目してみると、図5の様に、音声区間 (Speech Segment), 音楽区間 (Music Segment), 雑音区間 (Noise Segment), さらにはそれらが複数重なり合って形成される区間などが存在する。本稿では、音声、雑音が混在しない音楽区間がマルチメディアコンテンツに存在すると仮定した上で、その音楽区間を用いて、全体の音場の系 (伝達関数) を周波数的に推定する。

3.3 要素技術

システムの基礎要素として、リファレンスを用いた音源分離が必要となることはすでに述べた。これまで研究されてきた音源分離手法の多くはブラインド音源分離法などに代表されるように、混合音のみからの目的音抽出が主流である。本検討では、あらかじめ混合音に含まれる、抽出対象の音楽がリファレンスとして保持されている場合を前提条件とし、その音源分離法を検討する。音源分離では、複数の音声が存在するなど、相関性の高い音源を扱うことが想定されるが、リファレンス音を用いることにより、混合音のみの音源分離と比較して、除去精度の向上、処理の単純化や処理負荷の軽減が期待できる。

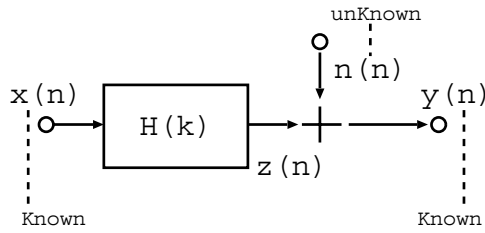


図6 伝達関数推定のための想定入出力モデル

4. 音源分離

4.1 クロススペクトル法

1983年, Carter, Knapp, Nuttallらは[9]において, クロススペクトル法と呼ばれる伝達関数の推定を行い, 同時に, 伝達系における因果性の線形性の評価尺度となるコヒーレンス関数の推定方法を開発した. 現在, クロススペクトル法は伝達関数推定において標準的手法として確立しており, 本稿においてもこの手法を用いることとする. まず, 単純なモデルとして図6のような1入力, 1出力で, 更に出力信号に対してノイズ信号が加えられるモデルを考える. 出力信号 $y(t)$ には, 入力信号 $x(t)$ とノイズ成分である $n(t)$ が含まれているとする. ここでサンプリング間隔が τ の離散的な表現を用いて出力信号を表現すると, 以下の式で与えられる.

$$y(n) = z(n) + n(n) \quad (1)$$

$$= h(n) * x(n) + n(n) \quad (2)$$

$h(n)$ は伝達系のインパルス応答を示しており, *は畳み込み演算を示している. ここで, 入力信号 $x(t)$ と出力信号 $y(t)$ に対し, N 点 DFT(Discrete Fourier Transform) を施すと式(2)は以下のように表せる.

$$Y(k) = H(k)X(k) + N(k) \quad (3)$$

離散スペクトルは次式で与えられる.

$$X(k) = \sum_{n=0}^{N-1} x(n) \exp(-j2\pi \frac{kn}{N}) \quad (4)$$

標準化信号 $x(n)$ と $y(n)$ を, 一つ一つが N 点からなる M 個のブロック $x_i(n)$ と $y_i(n)$, ($i = 1, 2, \dots, M$) に分割する. そのおのにおに N 点の離散的フーリエ変換を適用し, 得られたおのの離散的なスペクトルを $X_i(k)$ と $Y_i(k)$ と表す. k は離散的周波数を表す整数である. $X_i(k)$ と $Y_i(k)$ は次のように与えられる.

$$X_i(k) = \sum_{n=0}^{N-1} x_i(n) \exp(-j2\pi \frac{kn}{N}) \quad (5)$$

$$Y_i(k) = \sum_{n=0}^{N-1} y_i(n) \exp(-j2\pi \frac{kn}{N}) \quad (6)$$

ここで, $N_i(k)$ は i 番目のブロックの雑音系列 $n_i(n)$ のスペクトルである. i 番目のブロックに関して, 入力のスペクトル $X_i(k)$ から推定できる伝達系の応答 $Z_i(k) = H(k)X_i(k)$ と, 出力端で実際に観測された信号のスペクトル $Y_i(k)$ との差の成分のパワー α は式(7)で与えられる.

$$\alpha = E_i[|Y_i(k) - Z_i(k)|^2] \quad (7)$$

この平均パワー α を最小にするような $H(k)$ を求めることによって, 伝達関数 $H(k)$ の最適値が与えられる.

$$H(\hat{k}) = \frac{E_i[X_i^*(k)Y_i(k)]}{E_i[|X_i(k)|^2]} \quad (8)$$

さらに,

$$H(\hat{k}) = H(k) + \frac{E_i[X_i^*(k)N_i(k)]}{E_i[|X_i(k)|^2]} \quad (9)$$

となる. これより, 伝達系の出力信号 $y(n)$ に雑音 $n(n)$ が混入する場合でも, 雑音 $n(n)$ と $x(n)$ が無相関ならば, 加算平均回数 M の増加に伴って, 式(9)の右辺第2項目の雑音成分の平均振幅を相対的に $\frac{1}{\sqrt{M}}$ 倍に減少できる. 従って, M 回の振幅比で $10 \log_{10} M$ [dB] の信号雑音比 (SNR) の改善が得られる.

4.2 伝達関数推定

本章では提案システムに実装している音源分離手法について述べる. 具体的な音源分離処理について図8に示した. 提案システムにおいて, 伝達関数を推定するための手法であるクロススペクトル法では, リファレンス音楽信号と出力信号 $y(n)$ を用いている関係上, その推定精度を向上させるため, マルチメディアコンテンツの構成(図5)においても示したように純粋な音楽信号のみが存在する区間を用いて計算を行う. このリファレンス音楽信号を用いることは本検討の重要な前提条件である.

出力信号 $y(n)$ はインパルス応答 $h(n)$ と様々な要素を含む入力信号 $x(n) (= x_1(n) + x_2(n))$ から生成される. ここで, 図8に示した伝達関数 $H(k)$ は以下のように示すことができる.

$$H(k) = \frac{y(k) * x_2^*(k)}{x_2(k) * x_2^*(k)} \quad (10)$$

式(10)において, $y(k) * x_2^*(k)$ は $x_2(k)$ のパワースペクトルを示し, $y(k) * x_2^*(k)$ は $x_2(k)$ と $y(k)$ のクロススペクトルを示し, 伝達関数はその比によって与えられる. さらに, 仮想的に音場を通過した音楽信号 ($y'(n)$) を作り出すために, リファレンス音楽信号の周波数変換 $X_2(k)$ とすでに推定された伝達関数 $H(k)$ を用いて信号の合成を行う. 最終的に, 現実環境で観測された出力信号と仮想的に作り出された音楽信号の差分を取ることで, 目的の音楽信号を除去した信号を取り出すことができる. 次章において更に詳しく説明を加える.

4.3 信号除去処理

本稿の基本的な発想は, リファレンス信号 $x_2(n)$ を用いた音場の伝達関数推定により, 配信権が自分にはない音楽信号を分離・除去することである. 以下にその信号除去処理の具体的な処理過程に関して説明する. 推定した伝達関数を元に音楽信号の除去を行なう(図8). 推定した伝達関数と周波数変換された音楽信号の積によって, 仮想的にある音場を音楽信号が通過した時の信号を作り出す. 実際に出力音として検出されている出力信号と, この仮想的に与えられた音楽信号との差分により音楽情報の除去を行なうことができる. 以下に具体的な手順について述べる.

あらかじめ測定されている入力信号 $x(n)$ と, 出力信号 $y(n)$ を用いて式(10)より, 伝達関数 $H(k)$ を算出する. 同時に入力信号に含まれている音楽信号 $x_2(n)$ をリファレンス信号として保持していることが前提条件であるため, 求めた $H(k)$ を用いて入力信号が通過したものと同一音場を音楽信号 $x_2(n)$ が通過したときの出力信号を計算する.

$$X_2(k) = \sum_{n=0}^{N-1} x_2(n) e^{-j2\pi \frac{kn}{N}} \quad (11)$$

$$Y'(k) = X_2(k) * H(\hat{k}) \quad (12)$$

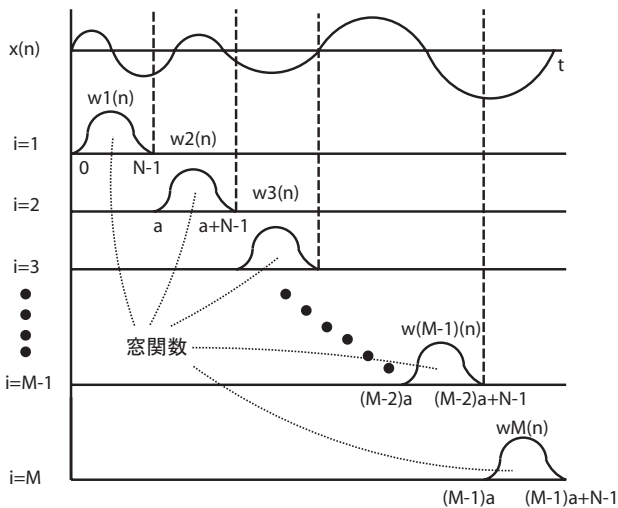


図7 クロススペクトル法

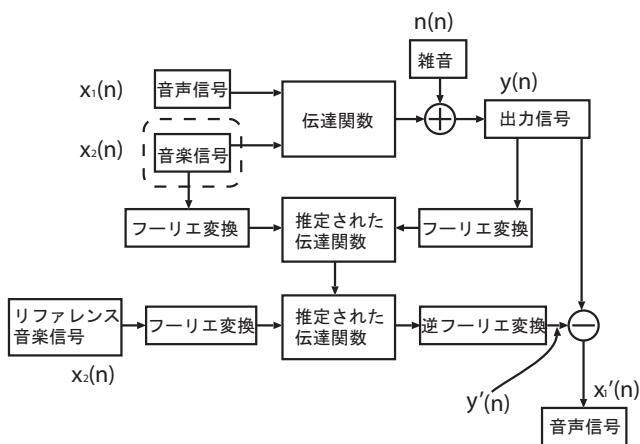


図8 提案する音源分離手法の処理プロセス

式 (12) により得られた $Y'(k)$ に逆フーリエ変換したものが、音楽信号が入力信号と同等の音場を通過したときの出力信号である。これより $y'(n)$ は

$$y'(n) = \frac{1}{N} \sum_{k=0}^{N-1} Y'(k) e^{-j\frac{2\pi}{N}kn} \quad (13)$$

$$f = \frac{k}{NT} \quad (14)$$

実際に観測した出力信号 $y(n)$ は、音声信号 $x_1(n)$ と音楽信号 $y_2(n)$ の混合信号を伝達関数が $H(k)$ で表現可能な音場に通じたときの信号である。この出力信号 $y(n)$ と $y'(n)$ の差分をとることにより、音場 $H(k)$ を通過した時の音声信号 $x'_1(n)$ を作り出すことが出来る。

$$x'_1(n) = y(n) - y'(n) \quad (15)$$

5. シミュレーション

システムの有効性を示すために、本提案手法を用いてシミュレーションを行う。実際に音楽信号、音声信号を入力して伝達関数の推定を行い、出力信号を観測した。

本稿では、シミュレーションに際し、音声信号、音楽信号、雑音信号、その他の全ての信号に関してサンプリング周波数 44.1kHz、量子化ビット 16Bit の信号を用いた。窓関数はハニング窓を用い、その信号長はインパルス応答の 2 倍を割り当てた。

これら全ての信号はシミュレーション結果の評価のために、独立して保持しているものとして進めた。

システムの有効性評価のために行った実験についての手順を示す。まず、クロススペクトル法を用いたインパルス応答の推定を行い、その結果を示した。図 12(左)、図 12(右)、図 13(左)、図 13(右) はそれぞれ、クロススペクトル法における加算平均回数が 1 回、4 回、8 回、32 回の場合の結果である。図 11 は理想的なインパルス応答の信号波形である。

続いて、推定したインパルス応答を用いてシステムの出力信号を観測した。本稿ではクロススペクトル法における加算平均回数が 8 回と 32 回のインパルス応答を用いてシステムの出力波形を観測した。図 14(左) は 8 回加算平均、図 14(右) は 32 回加算平均のインパルス応答を用いたときのシステム出力波形である。また、図 10 は理想的な残響を持った音声信号波形を示している。

クロススペクトル法では、処理プロセス内の加算平均回数の違いにより得られるインパルス応答の品質が異なる。シミュレーションにおいては 1 回、4 回、8 回、32 回の加算平均を行ったインパルス応答波形を示した。さらにその推定インパルス応答を用いて、システムの出力波形を観測した。32 回の加算平均後のインパルス応答を用いた場合、理想環境下での十分な除去精度が観測できた。

6. まとめと今後の課題

従来の著作権保護技術は、マルチメディアコンテンツ自体に電子透かしを埋め込む、コンテンツにメタデータを付加することによって著作物を保護するなどの検討が主流であった。しかし、コンテンツの配信者が一般ユーザであった場合、配信したいコンテンツに対してメタデータを付加し、コンテンツを登録するなど様々な手続きを行うことは困難であると考えられる。さらには、このような手続きがマルチメディアコンテンツの流通を妨げる恐れもある。

本検討では、コンテンツ配信を行う際、コンテンツに含まれる、配信権がユーザにない音楽信号に注目し、その音源分離の検討を行った。伝達関数の推定をクロススペクトル法により行うことで、入出力にノイズ成分が存在する場合でも音楽情報は除去出来ることが確認できた。本提案では、ユーザが最終的にその出力信号を利用して、再び著作権を持たない音楽を付加し、配信することを考えているため、出力信号の品質についてはさらに検討を進めていく必要がある。

出力音の品質を向上させるためにも、伝達関数を最適にするための入出力信号の推定など、入力信号、処理結果として得られる出力信号に関しての前後処理が必要であり検討すべき課題である。

さらに、今後、システムの実装を考えたいという検討を進める際、音楽信号同士の同期が問題として挙げられる。残響、信号の遅延などの状況下において、どのような処理が最適な伝達関数を導いてくれるのかなどの検討も同時に進めていきたい。

謝 辞

本研究を進めるにあたり、日頃から惜しみなく御指導して頂きました早稲田大学国際情報通信研究センター渡辺裕 教授、小館亮之 助教授、及川靖広 講師に深く感謝致します。また、実験ソースの提供を頂いた (株) シアター・テレビジョンに感謝いたします。

文 献

- [1] Qiong Liu, Reihaneh Safavi-Naini and Nicholas Paul Shappard, "Digital Rights Management for Content Distribution," Australian Information Security Workshop 2003(AISW2003), Adelaide, Australia, Conference in Research and Practice in Information Technol-

ogy, Vol.21.C, 2003

- [2] 阪本秀樹, "Content ID Forum の標準化動向," 映像情報メディア学会, Vol. 55 No. 3, pp.353-358, 2001
- [3] Steve Czerwinski, Richard Fromm and Todd Hodes, "Digital Music Distribution and Audio Watermarking,"
- [4] Oliver Hellmuth, Eric Allamanche, Jurgen Herre, Thorsten Kastner, Markus Cremer, Wolfgang Hirsch, "Advanced Audio Identification Using MPEG-7 Content Description," AES111th Convention, New York, NY, USA, 2001 September 21-24
- [5] Thorsten Kastner, Eric Allamanche, Jurgen Herre, Oliver Hellmuth, Markus Cremer, Holger Grossmann, "MPEG-7 Scalable Robust Audio Fingerprinting," AES112th Convention, Munich, Germany, 2002 May 10-13
- [6] P.Common, "Independent component analysis a new concept?," Signal Processing, vol.36, pp.287-314, 1994.
- [7] 中谷智広, 柏野邦夫, 奥乃博, "背景音楽つき音声に対する音響ストリームの分離," 情報処理学会, 音楽情報科学 19-12, (1997)
- [8] 猿渡洋, 西川剛樹, 荒木章子, 牧野昭二, "時間領域 ICA と周波数領域 ICA を併用した多段 ICA によるブラインド音源分離," 日本神経回路学会第 11 回全国大会 講演論文集, pp. 99-100, Sept. 2001.
- [9] Carter,G.C., Knapp,C.H. and Nuttall,A.H., "Eastimation of the magnitude-squared coherence function via overlapped fast Fourier transform processing," IEEE Transaction of audio and Electroacoustics, AU-21, 4, pp. 337 ~ 344(1983)
- [10] 中川聖一, "音声認識研究の動向," 信学論, j83-D-2 No.2 pp.433-457 (2000)
- [11] 金井浩, "音・振動のスペクトル解析," コロナ社 (1999)

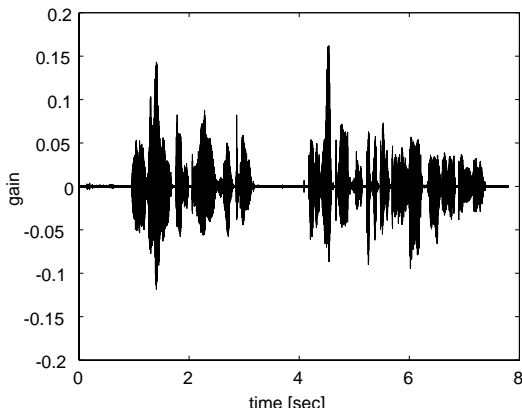


図 9 シミュレーションに用いた音声信号波形

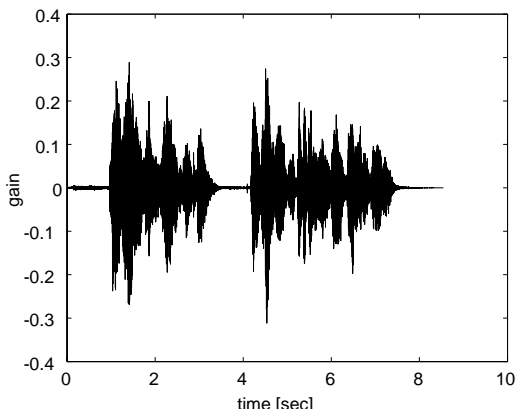


図 10 理想インパルス応答, 音声信号の畳み込み信号波形

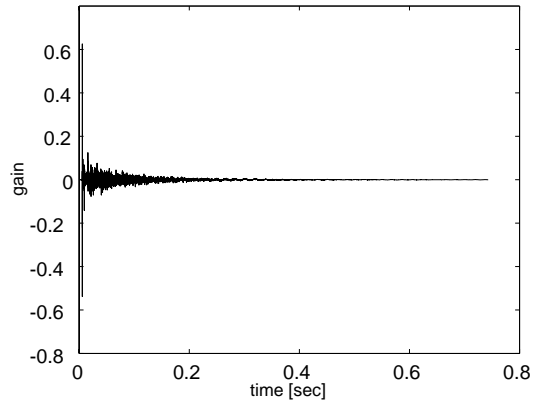


図 11 シミュレーションに用いたインパルス応答波形

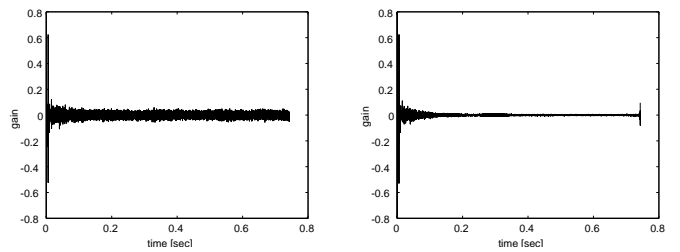


図 12 推定インパルス応答波形・1 回 (左), 4 回加算平均 (右)

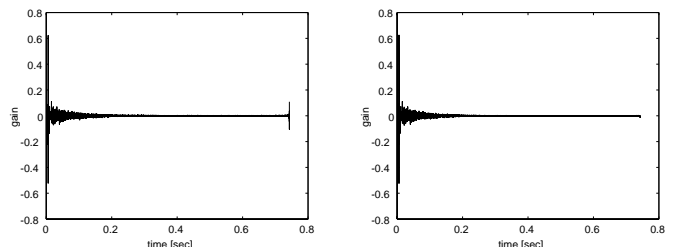


図 13 推定インパルス応答波形・8 回 (左), 32 回加算平均 (右)

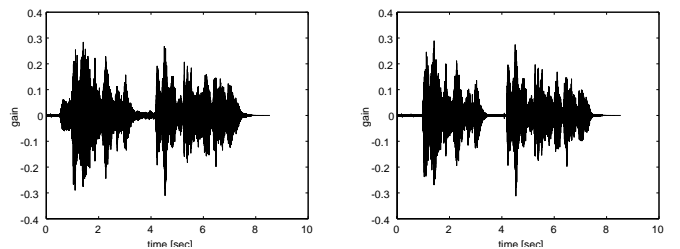


図 14 システムの出力信号波形・8 回 (左), 32 回加算平均 (右)