

平成 28 年度 数学協働プログラム・ワークショップ

ウェーブレット理論と工学への応用

OKU & ISM 2016 Workshop

Wavelet theory and its applications to engineering

主催：大阪教育大学，統計数理研究所

場所：大阪教育大学 天王寺キャンパス

日程：平成 28 年 11 月 22 日（火）13:20 － 18:00

平成 28 年 11 月 23 日（水） 9:30 － 15:00

開催趣旨

数学協働プログラム・ワークショップ「ウェーブレット理論と工学への応用」では、広い意味でウェーブレット解析によって解決できるかもしれないと期待できるトピックスに関して、講演者の方々に理論と工学的応用の現状、さらに解決すべき問題を解説していただき、その問題提起を端緒として参加者がディスカッションする形で、ウェーブレット解析が実際にどのように応用されているかをより深く理解することによって、新しい理論と応用への道が開かれることを目指します。

ウェーブレット理論と工学への応用

平成 28 年 11 月 22 日 – 23 日

大阪教育大学 天王寺キャンパス 西館 第 1 講義室

平成 28 年 11 月 22 日 (火) 13:20 – 18:00

13:20–13:30 開催の挨拶

13:30–14:30 章 忠, CHONG HUEI SHAN, 戸田 浩, 秋月 拓磨, 三宅 哲夫 (豊橋技術科学大学)

2次元複素数離散ウェーブレット変換による上咽頭粘膜病変部位の抽出 5

上咽頭粘膜病変は臨床的に特徴的な所見が乏しく、診断に苦慮する疾患である。これまでに、経験のある医者は内視鏡の画像により、上咽頭粘膜に特徴的な変化を観察し、診断を行っている。本研究では、客観的な診断方法を支援するために、2次元複素数離散ウェーブレット変換を用いて内視鏡検査により得られる画像から粘膜の病変部位を抽出し、数値化することを試みる。さらにこの手法の問題点と改善について議論する。

15:00–16:00 鈴木 俊夫, 善甫 啓一, 木下 保 (筑波大学)

ウェーブレットを用いたディストーションサウンドの特徴量抽出 19

ディストーションサウンドが用いられている音楽の耳コピの精度は、個人の能力や経験に大きく依存する。しかし、ディストーションフィルタはクリッピングを含む非線形変換であるため、Fourier変換を用いた解析は難しい。今回我々はウェーブレットを用いて、ディストーションサウンドの歪みの特徴量を提案する。さらに、提案手法を用いた耳コピへの応用についても紹介したい。

16:30–17:30 芦澤 恵太 (舞鶴高専), 森田 雅貴 (名城大学)

DCT とウェーブレット変換を縦横に組み合わせた直交変換をブロック毎に用いる適応的画像圧縮法 27

現在、普及している非可逆画像圧縮方式の多くは、ブロック単位の離散コサイン変換 DCT に基づいている。平成 25 年の本ワークショップにおいて、最も単純なウェーブレット変換であるハール変換を利用することで、DCT に由来する JPEG 標準の代表的な視覚的歪みを軽減するアイデアを紹介した。本講演では、その後の研究成果を報告すると共に、ハール変換を画像圧縮に用いた最近の研究についても報告する。

平成 28 年 11 月 23 日 (水) 9:30 – 15:00

9:30–10:30 井川 信子 (流通経済大学), 守本 晃, 芦野 隆一 (大阪教育大学)

他覚的聴力検査に用いる誘発脳波のウェーブレット解析についての考察 39

自分の意思で聞こえを応えることができない場合、他覚的聴力検査が実施される。その中でも耳から音刺激を与えて誘発される脳波を用いる検査がある。この検査は MRI などと比べると安価で行われる検査なので健診などでの活用が望まれるが、脳波解析の難しさのために検診レベルの実用化が

遅れている。計測時即時診断の実現を求める目的で、我々はこれまで聴性誘発脳波解析にウェーブレット解析を適用してきた。これまでの成果と課題についてまとめることで、問題点を明確にし解決策について議論する。

11:00–12:00 新井 康平 (佐賀大学)

ウェーブレットによるパラメトリック SAR 画像分類

53

地表面とマイクロ波とのインタラクションメカニズムを偏波情報を用いて解析することにより、画像分類する手法を紹介する。その際、偏波情報の分解による散乱メカニズムの解明を用いる方法も紹介する。さらに、これに離散ウェーブレット変換を考慮することにより、新たな情報抽出が可能になって分類精度向上が期待できることを示す。

昼食 12:00–13:30

13:30–14:30 滝口 孝志 (防衛大学校)

コンクリート建造物に対する超音波 CT 開発と窓関数の応用

71

コンクリート建造物に対する tomographic な非破壊検査技術は未開発であるというのが現状であるが、近年、三田紀行 (職業能力開発総合大学校) と滝口の協働により、コンクリート建造物に対する超音波 CT 開発のアイデアが提案されている。本講演では、コンクリート建造物に対する超音波 CT 開発のアイデアを紹介し、この問題に対する窓 Radon 変換の応用に関して議論する。

大阪教育大学 天王寺キャンパス 西館 第 1 講義室

〒 543-0054 大阪市天王寺区南河堀町 4-88 電話番号 (06)6775-6611

JR 天王寺駅、地下鉄天王寺駅、近鉄大阪阿部野橋駅下車、徒歩約 10 分。

JR 寺田町駅下車、徒歩 5 分。

<http://osaka-kyoiku.ac.jp/>

統計数理研究所 数学協働プログラム

<http://coop-math.ism.ac.jp/>

連絡先

守本晃 (大阪教育大学 情報科学)

e-mail: morimoto@cc.osaka-kyoiku.ac.jp

tel: 072-978-3665

<http://www.osaka-kyoiku.ac.jp/~morimoto/TENWS/ws2016HP/>

芦野 隆一 (大阪教育大学 数理科学)

e-mail: ashino@cc.osaka-kyoiku.ac.jp

tel: 072-978-3685

<http://www.osaka-kyoiku.ac.jp/~ashino>

無線 LAN 情報, カラー版予稿集

eduroam 使えます. ワークショップ専用無線 LAN は,

SSID: eventwifi (2.4GHz, 5GHz 共通)

PW: minamikawahori (南河堀; みなみかわほり)

です.

カラー版の予稿集は,

<http://www.osaka-kyoiku.ac.jp/~morimoto/WSPRO/>

からダウンロードできます.

謝辞

本ワークショップは, 大阪教育大学, 統計数理研究所 数学協働プログラム, 科研費 (C) 26400199 のサポートを受けています.

2次元複素数離散ウェーブレット変換による上咽頭 粘膜病変部位の抽出

章 忠* Chong Huei Shan[†] 戸田 浩* 秋月拓磨* 三宅哲夫*
* 豊橋技術科学大学工学部 [†] 豊橋技術科学大学大学院

概要. 上咽頭粘膜病変は咽喉頭酸逆流症による病状であるが、臨床的に特徴的な所見が乏しく、診断に苦慮する疾患である。これまでに、経験のある医者は内視鏡の画像により、上咽頭粘膜に特徴的な変化を観察し、診断を行っている。本研究では、客観的な診断方法を支援するために、2次元複素数離散ウェーブレット変換を用いて内視鏡検査により得られる画像から粘膜の病変部位を抽出し、数値化することを試みる。さらにこの手法の問題点と改善について議論する。

Extraction of Nasopharyngeal Mucosal Lesion Portions by the Two-dimensional Complex Discrete Wavelet Transform

Zhong Zhang* Chong Huei Shan[†] Hiroshi Toda*
Takuma Akiduki* Tetsuo Miyake*

*Toyohashi University of Technology

[†]Graduate School of Toyohashi University of Technology

Abstract. Nasopharyngeal mucosal lesion that caused by the laryngopharyngeal reflux disease (LPRD) is a kind of disease when the stomach acid is refluxed back to throat and affects the larynx and pharynx. And it is hard to diagnosis LPRD. In addition, all collected data are contaminated by noise that will affect the analysis result. In this study, We proposed that combining the 2D-CDWT with morphology in order to extract white dots from NBI of nasopharyngeal mucosa epithelial tissue.

1. はじめに

食生活の欧米化や加齢によって、胃酸逆流によって起こされる疾患群特に耳鼻咽喉領域の症状を訴えるもの LPRD(laryngopharyngeal reflux disease) [1], 日本語で「咽喉頭酸逆流症」と呼んでいることが多くなっている。日本では、1995 年咽喉頭酸逆流症に関する報告が最初であり、最近では関心が高まりつつある。しかし、日本での咽喉頭酸逆流症の症状は比較的に軽いため、胃酸逆流と様々な耳鼻咽喉科領域疾患、症状との関連はあると考えられているものの、確実な証明がなされていないのが現状である。また、咽喉頭異常感等の症状は様々な要因が関与しており、診断がそもそも複雑であることなどが考えられる。

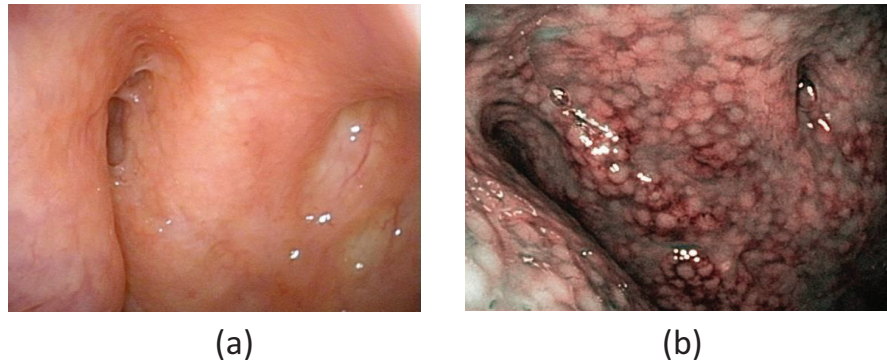


Fig. 1. 上咽頭粘膜の内視鏡画像の例

丹羽ら [2] は内視鏡を用いて LPRD 患者の上咽頭粘膜に特徴的な変化を認め、その粘膜変化の形状からうろこ雲パターン (mackerel cloud pattern) を定義した. 図 1 には上咽頭粘膜の内視鏡画像の例を示す. ただし, 図 1 の (a) は正常な上咽頭粘膜の内視鏡画像の例で, (b) は上咽頭粘膜の病変画像の例である. 図示のように, 正常な上咽頭粘膜では滑らかな表面であるが, 病変な上咽頭粘膜では粒上な模様があり, うろこ模様みたいなものが成形されている. それは, 咽喉頭酸逆流症で引き起った上咽頭粘膜炎症である. 本研究では, 上咽頭粘膜内視鏡の画像解析により上咽頭粘膜病変部位を抽出し, 医師に診断のための有用な情報を提示する目的とする.

従来の画像処理手法では, 製品画像のパターン情報等から欠陥を検出するフーリエ変換を使った手法やガボール関数等のフィルタリングによるエッジ検出手法を用いたものがある [3]. しかし, フーリエ変換を用いる場合は, テクスチャ等の特定パターンを持つものに用途が限られる. また, ガボール関数等のフィルタバンク用をいいる場合は, 計算量が大きく, 検査時間が長くなってしまう. そこで本研究では, ウェーブレット変換 (Wavelet Transform, WT) を用いた画像処理を検討する.

デジタル画像に WT を適用する場合, Mallat [4] が提案した多重解像度解析 (Multi Resolution Analysis, MRA) の高速アルゴリズムを用いた離散ウェーブレット変換 (Discrete Wavelet Transform, DWT) が多く用いられる. このため, 一般的なフィルタバンクと比べ, 計算量が少ないという利点を持つ. また, DWT は従来のフーリエ変換と比べ, 不規則なパターンや画像の解析や特徴検出が可能であるため, 適用可能な画像が多いという利点を持つ. さらに, MRA を用いているため, DoG 関数のようにパラメータを調節することなく, 各スケールの成分を抽出できる利点も持つ. しかし, MRA を用いた DWT では, 検出した信号の特徴の位置によって解析結果にばらつきが生じ, 時不変性を持たない. そのため, 従来, 頑健な画像処理が困難であった. この問題に対して, 戸田ら [5, 6] は完全シフト不変複素数離散ウェーブレット変換 (Perfect Translation Invariance Complex Discrete Wavelet Transform, CDWT) を提案している. CDWT は他にも kingsbury ら [7] や Selesnick ら [8] などが提案されたものもあるが, 本論文ではこれらを総称して, CDWT

とする．CDWT は信号の位相によらず，頑健な解析が可能であるという大きな利点を持つ．

一方，画像処理にはモルフォロジー演算という処理手法がある [9]．それは，画像が特定のパターンの集合で成形されるものと定義し，特殊の演算でパターン，または構造を抽出できる．本研究では，本研究は 2D-CDWT やモルフォロジー演算などを組み合わせ，咽喉頭酸逆流症による上咽頭粘膜病変の抽出を試みる．

2. 2次元複素数離散ウェーブレット変換とモルフォロジー演算の吟味

2.1 2次元複素数離散ウェーブレット変換の計算

従来の DWT や 2D-DWT には，シフト不変性の欠如と呼ばれる弱点，すなわち画像の特徴の位置によって変換結果が異なり，頑健な画像処理が困難となる弱点があった．戸田らの提案する CDWT は，Meyer の直交ウェーブレットを基礎に設計されており，完全シフト不変性を実現している [5]．CDWT は実数部と虚数部に分かれた，2つの直交ウェーブレットにより構成されるのが特徴である．すなわちスケーリング関数は，実数部のスケーリング関数 $\phi^R(x)$ ，虚数部のスケーリング関数 $\phi^I(x)$ があり，またマザーウェーブレット (Mother Wavelet, MW) も同じく，実数部の MW $\psi^R(x)$ ，虚数部の MW $\psi^I(x)$ がある．なお MRA の高速アルゴリズムに用いる実数部のローパス・ハイパスフィルタのフィルタ係数を $\{a_k^R\}$ ， $\{b_k^R\}$ ，また虚数部のそれらを $\{a_k^I\}$ ， $\{b_k^I\}$ とする．以上のような CDWT を 2次元に拡張した 2次元複素数離散ウェーブレット変換 (2D Complex Discrete Wavelet Transform, 2D-CDWT) について述べる．

2D-CDWT では，スケーリング関数と MW が，それぞれ実数部・虚数部を持ち，これらを用いて 2次元の信号 $f(x, y)$ を式 (2.1) を用いて展開する．ここで，RR から II の内，RR, RI は以下の式 (2.2)，(2.3) で示される．

$$(2.1) \quad f(x, y) = RR(x, y) + RI(x, y) + IR(x, y) + II(x, y)$$

$$(2.2) \quad \begin{aligned} RR(x, y) = & \sum_{k_x, k_y} c_{J, k_x, k_y}^{RR} \phi_{J, k_x}^R(x) \phi_{J, k_y}^R(y) + \sum_{j=J}^{-1} \sum_{k_x, k_y} d_{j, k_x, k_y}^{LH, RR} \phi_{j, k_x}^R(x) \psi_{j, k_y}^R(y) \\ & + \sum_{j=J}^{-1} \sum_{k_x, k_y} d_{j, k_x, k_y}^{HL, RR} \psi_{j, k_x}^R(x) \phi_{j, k_y}^R(y) + \sum_{j=J}^{-1} \sum_{k_x, k_y} d_{j, k_x, k_y}^{HH, RR} \psi_{j, k_x}^R(x) \psi_{j, k_y}^R(y) \end{aligned}$$

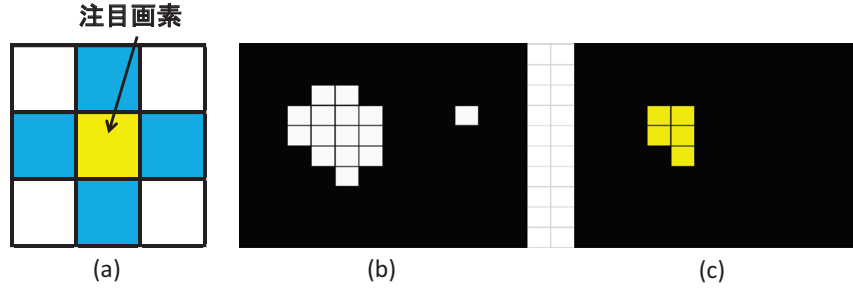


Fig. 2. 縮小処理の例

$$\begin{aligned}
 RI(x, y) = & \sum_{k_x, k_y} c_{J, k_x, k_y}^{RI} \phi_{J, k_x}^R(x) \phi_{J, k_y}^I(y) + \sum_{j=J}^{-1} \sum_{k_x, k_y} d_{j, k_x, k_y}^{LH, RI} \phi_{j, k_x}^R(x) \psi_{j, k_y}^I(y) \\
 (2.3) \quad & + \sum_{j=J}^{-1} \sum_{k_x, k_y} d_{j, k_x, k_y}^{HL, RI} \psi_{j, k_x}^R(x) \phi_{j, k_y}^I(y) + \sum_{j=J}^{-1} \sum_{k_x, k_y} d_{j, k_x, k_y}^{HH, RI} \psi_{j, k_x}^R(x) \psi_{j, k_y}^I(y)
 \end{aligned}$$

式 (2.2), (2.3) 中の c_{k_x, k_y}^{RR} , c_{k_x, k_y}^{RI} はスケーリング係数であり, スケーリング関数と解析信号の補間処理および再帰的な分解アルゴリズムによって計算される. ここで, J は分解レベルを示す. c_{k_x, k_y}^{RR} は, 信号の低周波成分を表現するための係数である. なお IR , II の他のパートにおいても同様である. 補間処理の計算等は文献 [10] を参照されたい. 次に, $d^{LH, RR}$, $d^{HL, RR}$, $d^{HH, RR}$ 等は, ウェーブレット係数と呼ばれ, 各周波数帯域の成分に相当する. 各ウェーブレット係数の計算は, 補間処理後の入力信号に対し, ローパス・ハイパスフィルタを適用する. そして, 計算結果をダウンサンプリングし, 各ウェーブレット係数を得る. また, 各レベルの $\psi_{j, k_x}^R(x)$, $\psi_{j, k_y}^I(y)$ 等のウェーブレットには, 空間 (時間) 領域において, 拡大・縮小の関係がある. 各ウェーブレット係数の計算方法およびレベル間のウェーブレットの関係の詳細についても, 文献 [10] を参照されたい. 2D-CDWT では, 低周波成分に相当する c_{k_x, k_y}^{RR} から c_{k_x, k_y}^{II} の 4 つのスケーリング係数と, 各パート, 各分解レベルの $d^{LH, RR}$, $d^{HL, RR}$, $d^{HH, RR}$ 等を出力として得る.

2.2 モルフォロジー演算 [9]

モルフォロジー演算は, 対象とする画像内部に構造要素と呼ばれる小さな図形を当てはめ, それにより画像内部の対象となる部分を抽出する手法である. モルフォロジー演算は幾つかの基本処理により行なわれ, それらの例を以下に示す.

2.2.1 縮小処理 (Erosion)

縮小処理はモルフォロジー演算の基本処理のひとつである. その処理は画像の中に 2 個連結した領域を分離し, 細かい部分を削って内部の穴を拡大する. もと画像 A から構造化要素 B を引くことで, モルフォロジー変換処理である. 図 2 には縮小処理の例を示す.

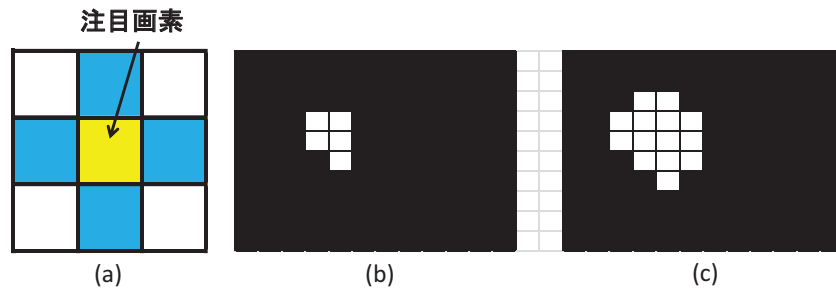


Fig. 3. 膨張処理の例を示す．ただし，図の (a) は張処理の例

ただし，図の (a) は縮小処理に使用する構造 B で，(b) は処理対象画像 A ，(c) は処理の結果である．図 2 の (a) に示す構造 B には，黄色部分は注目画素であり，青色部分は判断画素を表す．縮小処理は構造 B を対象画像 A に走査させ，青色部分に対応する画像 A の要素が全部画素 1 のとき，注目画素が 1 に残し．それ以外の場合に注目画素を 0 にする．この演算は次式で表すことができる．

$$(2.4) \quad A \ominus B = \{Z \in E : x - b \in A, b^V B\}$$

2.2.2 膨張処理 (Dilation)

モルフォロジー演算の基本処理の一つで，膨張処理を実施する．膨張処理では，画像の小さな穴等がふさがれ，分離していたものを繋ぐ．膨張処理は以下の式で表すことができる．

$$(2.5) \quad A \oplus B = \{Z \in E : z = a + b, a \in A, b \in B\}$$

ただし， a は元の画像 A の要素， b は構造化要素 B の要素である．式 (2.5) の処理結果は元画像 A と構造化要素 B の要素理論和をとったものである．本研究では膨張処理を使い，うろこ模様のサイズを調整する．図 3 には処理の例を示す．ただし，図の (a) は構造 B の例，(b) は処理対象画像 A ，(c) は処理の結果である．図の (a) の黄色部分は注目画素であり，青色部分は判断画素を表す．膨張処理は構造 B をもと画像 A に走査させ，青色部分の画素は 1 個また以上が 1 のとき，注目画素が 1 にする．それ以外の場合，注目画素を 0 にする．

2.2.3 オープニング処理 (Opening) とクロージング処理 (Closing)

オープニング処理は縮小処理と膨張処理 (3.2 章) を組み合わせ，一つステップとして使用するものである．ステップの内容は最初に縮小処理を行い，次に膨張処理を実施する．それは図形の凸部や孤立した小さな領域のみを除去する作用を持っている．クロージング処理は膨張の次に縮小処理を行い，図形の凹部や穴のみを除去する作用をもつ．図 4 にオープニング処理とクロージング処理の例を示す．ただし，(a) はオープニング処理，(b) はクロージング処理の例である．

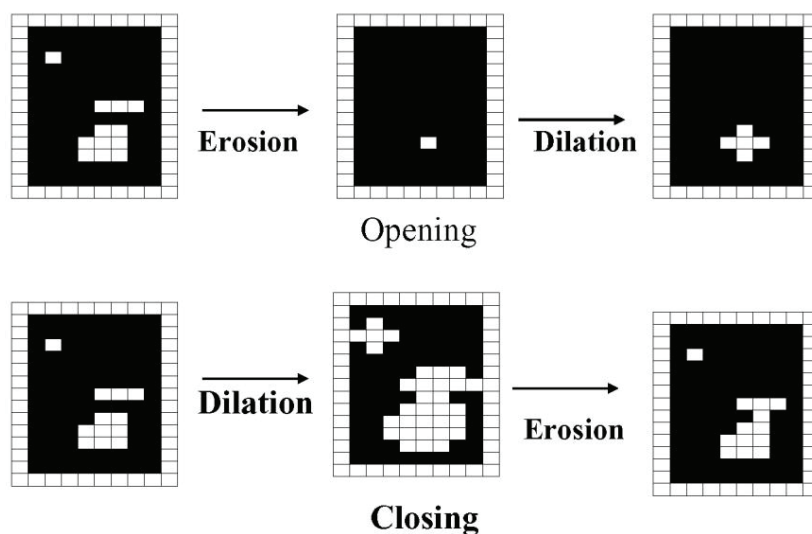


Fig. 4. オープニング処理とクロージング処理の例

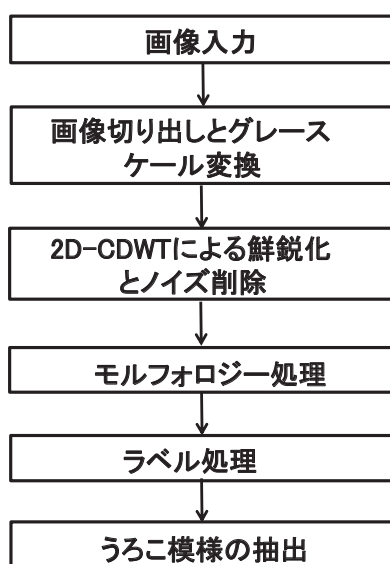


Fig. 5. 処理の流れ

3. 上咽頭粘膜病変部位の抽出

本節ではウェーブレット変換とモルフォロジー処理などによる病変部位の抽出について述べる．図 5 に示すのは処理の流れである．それは画像の切り出しとグレースケール変換，2D-CDWT による画像鮮鋭とノイズ削除，モルフォロジー処理，ラベル処理，うろこ模様の抽出からなり，次の節からそれぞれ説明する．

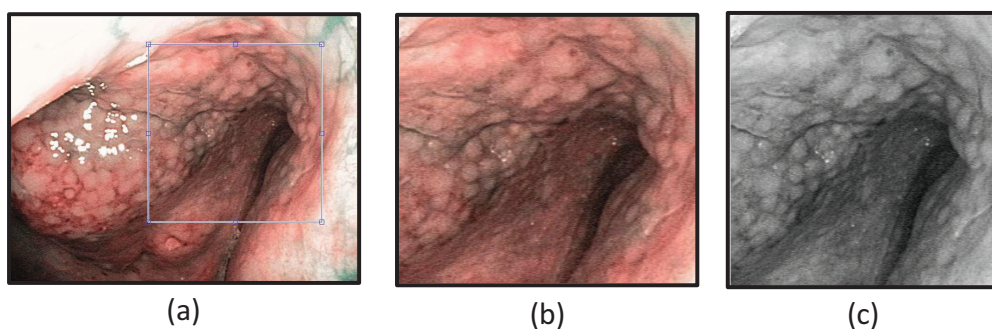


Fig. 6. 内視鏡画像およびグレースケース画像の例

3.1 画像切り出しとグレースケール変換

内視鏡画像では反射光が多く、そのまま処理すると、反射光はうろこ模様として抽出されてしまうため、反射光を避けながら画像を切り出す方法を考えた。本研究では、まずすべての内視鏡画像を同一な条件で処理するために、予め 320×320 画像が切り出せる窓を準備し、反射光がないところまで、手動で窓を移動して画像を切り出す。また、画像を人間の目で見ると、色付きの画像のほうが見えやすい。しかし、パソコンで画像処理を行うときに、計算量が多く、計算時間もかかる。それと比較して、画像をグレースケールにしたほうが計算しやすい。従って、次には切り出された画像を次式でグレースケールへ変換する。ただし、式中の R は画像の赤色画素、 G は緑色画素、 B は青色画素を表している。

$$(3.1) \quad \text{Image}(x, y) = 0.299 \times R(x, y) + 0.587 \times G(x, y) + 0.144 \times B(x, y)$$

図 6 には内視鏡画像および切り出された画像のグレースケール変換結果を示している。ただし、図 6 の (a) は内視鏡の元画像で、(b) は切り出された画像、(c) はグレースケール変換された画像の例である。Fig.6.2 では Fig.6.1 (b) をグレースケール変換にした出力画像を示す。

3.2 2D-CDWT による鮮鋭化とノイズ除去

図 7 に示すのは、グレースケール変換した画像を読み込み、2D-CDWT を実施し、鮮鋭化とノイズ除去を行った結果の例である。ただし、図 7 の (a) は 2D-CDWT により画像を分解した結果で、(b) は分解された画像を鮮鋭化した後、再構成された画像 (c) は鮮鋭化とノイズ除去を行ってから、画像を再構成した結果である。ウェーブレット変換は、2D-CDWT に使用したマザーウェーブレット (MW) は複素数ウェーブレットの一種である $m=3,4$ RI-Spline ウェーブレットであり、画像分解処理は Level-4 まで実施した。また、画像鮮鋭化処理はエッジを強調するためによく使用される手法である。本研究では、うろこ模様の輪郭を強調するために、2D-CDWT で分解された画像に対して、Level-2 での

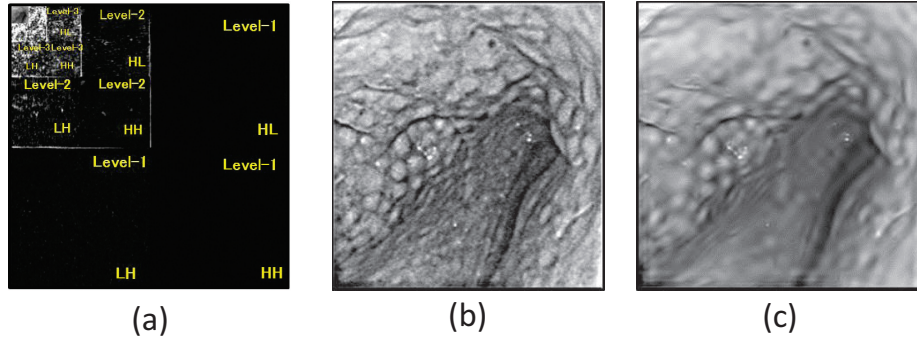


Fig. 7. 2D-CDWT による画像処理の例

ウェーブレット係数を 2 倍にし, Level-3 と Level-4 のウェーブレット係数をそれぞれ 3 倍にした. これは, 図 7 の (a) に示すように, Level-1 にはうろこ模様の情報がほとんど見られなく, Level-2, Level-3 と Level-4 にはうろこ模様の情報が明確に確認できたためである. 図 7 の (b) から分かるように, 画像鮮鋭化処理の結果はうろこ模様の輪郭を強調した. しかし, 不要なノイズでも強調されてしまうことも確認した. それは上咽頭粘膜の表面が粗くなったことからである.

そこで, 本研究はノイズ除去にウェーブレット縮退法を使用した. ウェーブレット縮退法は Donoho [11] により提案された手法で, ウェーブレット変換により分解された各ウェーブレット係数に対して目的信号とノイズの選別を行い, 得られた新たなウェーブレット係数を再構成して目的信号を抽出する手法である. 以下に Soft thresholding と呼ばれるウェーブレット縮退法の処理式を示す.

$$(3.2) \quad |\hat{d}_{j,n}| = \begin{cases} |d_{j,x,y}| - \lambda, & |d_{j,x,y}| > \lambda \\ 0, & |d_{j,x,y}| \leq \lambda \end{cases}$$

ただし, $|d_{j,x,y}|$ は各レベル j でのウェーブレット係数の絶対値 (ノルム), 添え字 (x,y) はデータの座標, λ は閾値, $|\hat{d}_{j,x,y}|$ は分離された目的信号のウェーブレット係数の絶対値である. また, 閾値 λ は次式で与えられる.

$$(3.3) \quad \lambda = k\sigma$$

ここで, 式中の σ はウェーブレット係数の標準偏差とし, 次式で求める.

$$(3.4) \quad \sigma = \frac{\text{median}(|d_{j,x,y}| - \text{median}(|d_{j,x,y}|))}{0.6745}$$

ただし, $\text{median}(|d_{j,x,y}|)$ は $|d_{j,x,y}|$ を大きさの順に並べた中央値である.

式 (3.2) によれば, 目的信号の振幅が閾値より大きい場合には, 閾値以下の信号を解析信号から減算し, またはその部位をゼロ化することによって, 振幅の小さい“ノイズ”を除去することができる. このような処理は, 個々のウェーブレット係数に対して行い, 信

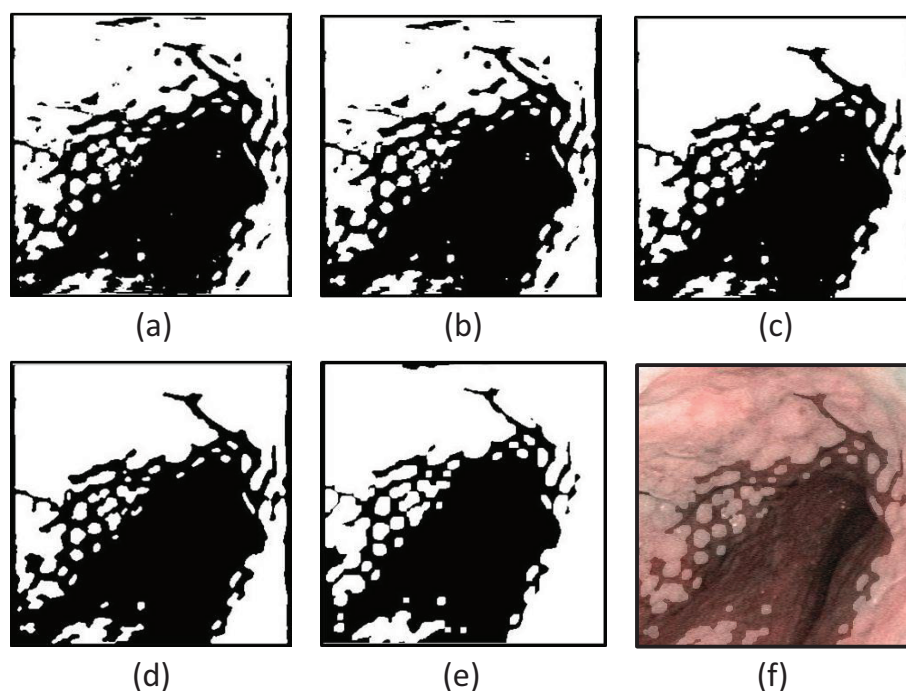


Fig. 8. モルフォロジー処理の例

号の全体的な形状を損なうことなく，細かな信号分離が可能になる．また，式 (3.3) 中の k について，Level-1 はうろこ模様の情報がないたため， $k = 4.0$ にした．それから，Level-2 と Level-3 は $k = 1.0$ にした．これは Level-2 と Level-3 にうろこ模様の情報が入っており，その情報を残すためである．Level-4 は低い周波数の帯域であるため， k を高すぎると，うろこ模様の情報も消えてしまうので， $k=0.3$ にした．また，これらの k の値は正常の上咽頭粘膜も適用するように決めた値である．図 7 の (b) と (c) を比較してわかるように，ノイズ除去後の画像では比較的滑らかで，うろこ模様の情報は目安くなった．

3.3 モルフォロジー処理

図 8 には図 7 の (c) に対してモルフォロジー処理を行った結果である．ただし，図 8 の (a) は 7 の (c) を二値化にした結果で，(b) は二値化の画像 (a) に対するオープニング処理，(c) は画像 (b) に対してクロージング処理，(d) は画像 (c) に対して Majority 処理，(e) は画像 (d) に対して膨張処理を行った結果である．最後の (f) は以上のモルフォロジー処理の結果と図 6 の (b) に示した元画像を重ねて表した図である．

図 8 の (a) に示すように，2 値化処理を行った画像にはうろこ模様の中に小さい黒点があり，上咽頭の輪郭では小さい白点もたくさんあった．これをオープニング処理を実施した結果は図 8 の (b) に示す．図示のように，小さい白点は消されましたが，やや大きい白点（うろこではないもの）はまだ消されていないことも確認した．そして図 (b) に対してクロージング処理を行った結果は (c) となり，白色領域内の黒色部分を埋めて分断された白

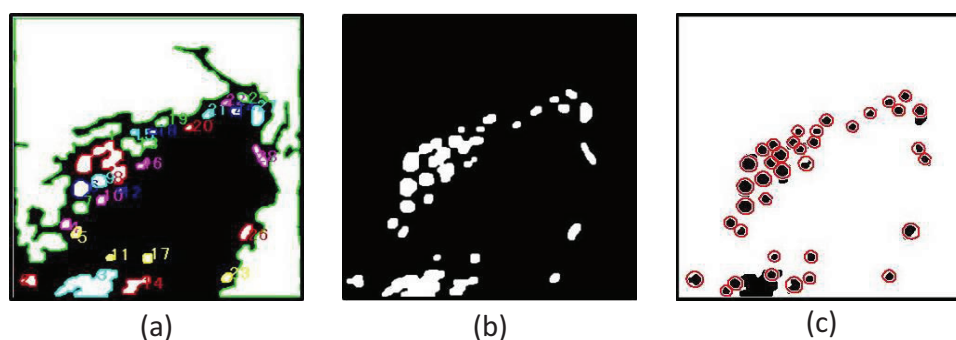


Fig. 9. ラベリング処理の例

色領を連結した。しかし、図 (c) を見ると、うろこ模様はギザギザの状態である。うろこ模様の形を整えるために、Majority 処理を行った。その結果は図 (d) であり、うろこ模様の形を一層円滑になり、理想な形状に近づいた。さらに、うろこ模様のサイズを整えるために、画像 (d) に対して膨張処理を行い、モルフォロジー処理の結果として画像 (e) に示す。ここで、それを元画像に重ねる図 (f) を見ると、うろこ模様のサイズが元画像のうろこ模様のサイズと近く、うろこ模様の抽出が成功であると言える。

3.4 ラベリング処理

ラベリング処理はデジタル画像処理によく用いる手法で、2 値化画像を構成する連結部分にラベル (番号) を付けることにより画像を構成する画像要素の分類や抽出などに使用される。本節では、ラベリング処理を使用し、各領域を番号付け、指定した範囲以内の領域における画像要素だけ残して抽出する。抽出された領域内の画像要素のパターンはうろこ模様として考え、その数や面積を計算するに使用することができる。

図 9 に示すのはラベリング処理の例である。ただし図 9 の (a) は仮ラベリング処理の例で、(b) はラベリング処理により抽出した領域の例、(c) は等量円でマックしたうろこ模様の例である。ラベリング処理は、まず各領域を分け、ラベリング (番号付け) をする仮ラベリング処理を行う。次にラベリングされた各領域内の面積をカウントし、ある範囲以内の画素数を有する領域だけラベルを残す。この処理はラベリングされた各領域から小さなノイズや大きな模様などを排除し、うろこ模様の領域だけを残すことができる。図 9 の (b) は (a) のラベリング処理の結果により残されたうろこ模様の部分を反転して得られた画像である。図示のように、うろこ模様の部分を抽出できたことを言える。

最後に、専門医の知見に基づいて抽出されたうろこ模様の大きさを等量円で評価することを試みる。本研究では、うろこ模様の等量円を抽出するには円検出のハフ変換を使用した。円検出ハフ変換はハフ変換の一つであり、直線を検出するハフ変換を円検出へ拡張した処理法である。円検出ハフ変換は、ノイズが混入した画像や検出対象物が明瞭に映っていない画から円を安定に抽出できる特徴を持ち、図 9 の (b) に示すような不規則な図

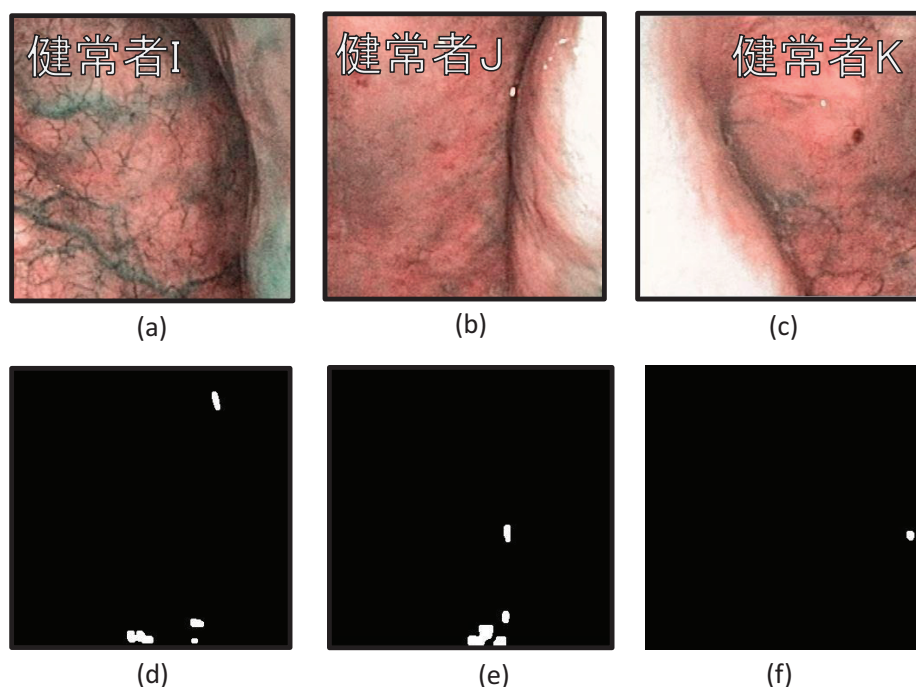


Fig. 10. 健常者の内視鏡画像の処理結果の例

形や欠損円弧が存在するうろこ模様から等量円を算出することに期待ができる．図 9 の (c) に示すのは 円検出ハフ変換を用いて処理した結果である．図示のように，等量半径 $6 < R < 12[\text{pixel}]$ 以内のうろこ模様を抽出することができた．

4. 検証実験結果と考察

図 10 に示すのは健常者の内視鏡画像の処理結果の例である．ただし，図 10 の (a), (b), (c) は健常者 I, J, K の内視鏡画で，(d), (e), (f) はそれらの処理結果である．図 10 に示した健常者の内視鏡画像の中に，特に (a) 健常者 I の場合に太い血管と網状の血管が見られている．その解析結果は (d) となり，血管の一部はうろこ模様として抽出してしまった．これに対して，健常者 J と K の場合には，それらの解析結果が (e) と (f) となり，少々の誤検出があるが，おおむねに良好な解析結果が得られた．ここでは，健常者の滑らかな表面に網状の血管が存在する場合には，それらをうろこ模様と区別できる手法の開発が今後の課題となる．

図 11 に示すのは患者の内視鏡画像からのうろこ模様検出例である．ただし，図 11 の (a), (b), (c) は健患者 C, D, E の内視鏡画で，(d), (e), (f) はそれらの処理結果である．図 11 に示す検証用内視鏡画像は本手法の基準画像となる図 6 の (b) と比較し，画像が少々暗く，うろこ模様も少しくらい緑色の粘膜で現れている．これらの画像の処理した結果は，うろこ模様が概ねに抽出できた．しかし，反射光（白色の部位）はうろこ模様として抽出してしまうことが確認されました．また，大きいサイズうろこ模様（カメラとの距離と近

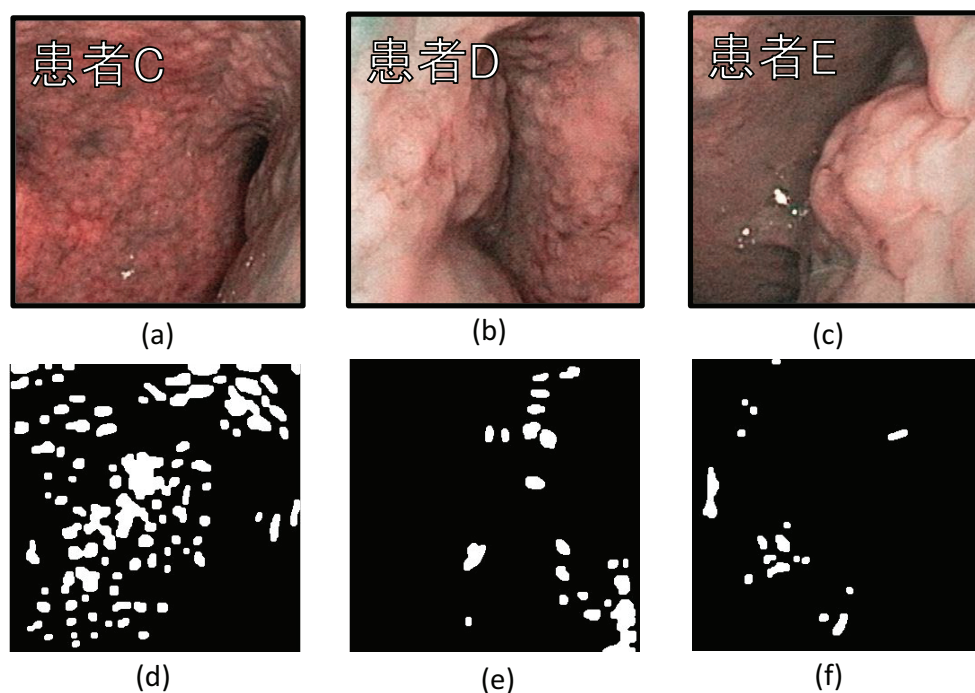


Fig. 11. 患者の内視鏡画像からのうろこ模様検出例

いうろこ模様)は抽出することができなかった。

本研究ではパラメータを設定するために、医師にすべての画像を同じ照明の設定で内視鏡カメラで撮影することを依頼した。しかし、患者の粘膜の状態は人により違うので、同じ照明でも得られた内視鏡画像の明暗度が違う。また上咽頭粘膜の画像はすべてピンク色ではないことが分かった。例えば、粘膜で血管や筋が多い方の内視鏡画像は同じ照明で撮影しても画像は比較的暗い。そのため、本研究の結果では暗い背景の画像ではうろこ模様を抽出できなかった。今後の課題では提案手法を実施する前に画像の輝度を合わせたら、暗い画像でもうろこ模様が抽出できると考えられる。

うろこ模様の輪郭はとても薄い(うろこ模様とほぼ同じ色)部分ではうろこ模様が抽出できなかった。それはうろこ模様の輪郭とうろこ模様の情報は 2D-CDWT で同じレベルに属するので鮮鋭化処理を行っても輪郭は強くならなかった。輪郭部分とうろこ模様の凸部分も一緒に鮮鋭化されたことが原因だと考えられる。将来にうろこ模様の色情報を把握し、うろこの輪郭の色の分布とうろこ模様凸部分の色の分布を離れるなら、以上の問題点解決できると考えられる。

5. まとめ

上咽頭粘膜病変は咽喉頭酸逆流症による病状であるが、臨床的に特徴的な所見が乏しく、診断に苦慮する疾患である。専門医の知見により、内視鏡の画像からの上咽頭粘膜病変の特徴として、正常な上咽頭粘膜の表面が滑らかであるに対し、上咽頭粘膜病変の粘膜

表面は炎症によるうろこ模様が現れることを挙げられる．本稿では，画像解析を用いて咽喉頭酸逆流症の上咽頭粘膜病変に見られる特徴的なうろこ模様パターンの解析手法を提案した．

本手法では，まず反射光の影響を避けるために，入力画像から病変部位を切り出し，グレースケール画像に変換する．次に切り出された画像について，2D-CDWTを行う．更に，病変部位を抽出しやすくするために，鮮鋭化処理とノイズ除去処理を行う．処理された画像に対してモルフォロジー処理を行い，病変部位を領域ごとに分割する．分割した領域にラベル処理より番号をつける．そして，番号付けられた領域について，ある大きさの範囲以内の領域だけ残し，それ以外の領域を削除する．最後に，残した領域から，うろこを抽出する．

提案手法により処理した結果は，内視鏡の画像からはうろこ模様が概ねに抽出することができた．しかし，うろこの輪郭の色とうろこ模様の色は近い場合にはすべてのうろこ模様を抽出することができなかった．それから正常な上咽頭粘膜ならもうろこ模様が少し抽出した．これは血管や筋状がある場合にそれらをうろこ模様として抽出してしまう問題点がある．今後の課題として，医師とともに解析結果を分析し，問題点を改善する方法を検討して医療支援システムの実現を目指す．

謝辞

本研究の一部は科研費基盤 (C)(課題番号：26420387) の補助を受けたことを付記し，謝意を表します．

参考文献

- [1] 渡辺雄介, 咽喉頭酸逆流症について, 日本小児耳鼻咽喉学会, 35(3):pp185-188, 2014.
- [2] Niwa H, Horihata S, Makiyama Y: Applicability of Narrow Band Image (NBI) to diagnose Laryngopharyngeal reflux disease (LPRD). 116th Annual Meeting of the American Academy of Otolaryngology Head and Neck Surgery (AAO-HNS), Washington DC, USA, 2012.9
- [3] 中野宏毅, 吉田靖夫, 藤田 和弘, ガボール関数を用いたカラー液晶ディスプレイのマクロ欠陥検査支援, 電子情報通信学会論文誌. D-II, 情報・システム, II-情報処理, Vol. J80-D-2, No. 3 (1997), pp.734-744.
- [4] Mallat, S. G., A Theory for Multiresolution Signal Decomposition The Wavelet Representation, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.11, No.7(1989), pp.674-693.
- [5] 戸田浩, 章忠, 完全シフト不変性を実現する複素数離散ウェーブレット変換, 信号処

- 理, Vol.12, No.2(2008), pp.156-166.
- [6] Toda H., Zhang Z., Perfect Translation Invariance with a Wide Range of Shapes of Hilbert Transform Pairs of Wavelet Bases, Int. J. Wavelets Multiresolution. Inf. Processing, Vol. 8, No. 4 (2010), pp.501-520.
- [7] Kingsbury, N. G., Image processing with complex wavelet(1999), pp.24-25, Philosophical Transactions of Royal Society London A.
- [8] Selesnick, W. I., The Design of Approximate Hilbert Transform Pairs of Wavelet Bases, IEEE Transactions on Signal Processing, Vol.50, No.5(2002), pp.1144-1152.
- [9] 石井明, 斉藤文彦, マシンビジョン, コロナ社, 2008 年 7 月 3 日初版.
- [10] 加藤毅, 章忠, 戸田浩, 今村孝, 三宅哲夫, 2 次元複素数離散ウェーブレット変換の方向選択性およびその半導体ウェーハの欠陥検査への応用, 日本機械学会論文集 C 編, Vol.79, No.808(2013), pp.4901-4916.
- [11] Donoho, D. L., De-noising by soft-thresholding, IEEE Transactions on Information Theory, Vol.41, No.3 (1995), pp.613-627.

章 忠 (豊橋技術科学大学工学部)

〒441-8580 豊橋市天伯町雲雀ヶ丘 1-1

E-mail: zhang@is.me.tut.ac.jp

Chong Huei Shan(豊橋技術科学大学大学院学生 (現在三菱自動車))

戸田 浩 (豊橋技術科学大学工学部)

〒441-8580 豊橋市天伯町雲雀ヶ丘 1-1

E-mail: pxt00134@nifty.com

秋月拓磨 (豊橋技術科学大学工学部)

〒441-8580 豊橋市天伯町雲雀ヶ丘 1-1

E-mail: akiduki@is.me.tut.ac.jp

三宅哲夫 (豊橋技術科学大学工学部)

〒441-8580 豊橋市天伯町雲雀ヶ丘 1-1

E-mail: miyake@is.me.tut.ac.jp

ウェーブレットを用いた ディストーションサウンドの特徴量抽出 Feature Extraction of Distortion Sounds Based on the Wavelet Transform

鈴木 俊夫¹, 善甫 啓一², 木下 保³
Toshio Suzuki, Kenichi Zempo and Tamotu Kinoshita

1. 筑波大学数理物質科学研究科, 2. 筑波大学システム情報系, 3. 筑波大学数理物質系
〒305-8577 茨城県つくば市天王台 1-1-1
1. Graduate School of Pure and Applied Sciences, University of Tsukuba
2. Faculty of Engineering, Systems and Information, University of Tsukuba.
3. Faculty of Pure and Applied Science, University of Tsukuba
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8577 Japan

Abstract

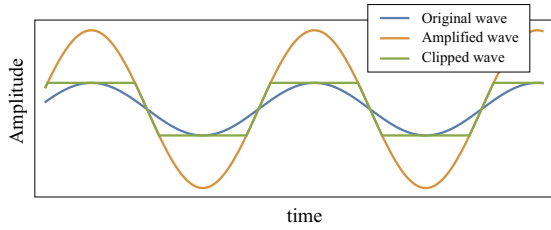
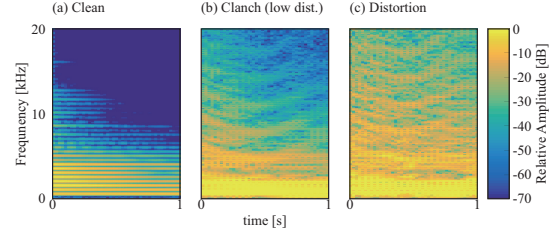
ディストーションサウンドが用いられている音楽の耳コピの精度は、個人の能力や経験に大きく依存する。しかし、ディストーションフィルタはクリッピングを含む非線形変換であるため、Fourier 変換を用いた解析は難しい。今回我々はウェーブレットを用いて、ディストーションサウンドの歪み量を提案する。さらに、提案手法を用いた耳コピへの応用についても紹介したい。

Dictation on the sound design of the distortion guitar depends on one's experiences and skills. Since the distortion effector includes the clipping process, which is nonlinear transform, it is difficult to analyze by the Fourier method. In this reason, we propose the method to analyze the distortion sounds with wavelets. Furthermore, we shall introduce the technique to dictate on the sound design.

1 はじめに

1.1 ディストーションフィルタとは

ポップやロック等のジャンルに分けられるような昨今の音楽では、必ずと言っていいほど、エレキギターを始めとしたデジタル信号を用いた楽器が使われている。そこでは音色を変えるために、エフェクタと呼ばれるフィルタを用いて、電気信号に変化を加えている。エフェクタにもディストーションエフェクタ、コーラスエフェクタ、コンプレッサエフェクタなどの様々な種類が存在し、演奏者はそれぞれのエフェクタを駆使しながら、好みの音色を構成していく。その中でも、歪みの一種であるディストーションエフェクタは必ずと言っていいほど用いられる、ギタリストには最も馴染みのあるエフェクタである。

Figure 1: $f(t)$ と $\check{f}(t)$, $\tilde{f}(t)$ の概形Figure 2: $f(t)$ と $\check{f}(t)$, $\tilde{f}(t)$ のスペクトログラム

音楽を演奏する際、耳コピという手段が頻繁に用いられる。これは音楽を聞き、演奏を再現したり楽譜に起こすというものである。この耳コピは、個人の技量や経験によるところが大きい。音高やリズムだけではなく、どの種類のエフェクタがどの程度かかっているのか等も耳コピで再現する際には、Fourier 変換を用いた解析は有効ではない [1]。耳コピによる演奏再現の 1 つの指標として、今回我々はディストーションの特徴量について提案する。

ディストーションエフェクタは次の処理を行っている：

ギターの信号 $f(t)$ が与えられたとする。以下、 $f(t)$ は正規化されている、即ち $\max |f(t)| = 1$ とする。ディストーションは主に 2 つの操作を行っている。ひとつ目は振幅の増加である。すなわち信号 $f(t)$ の振幅を大きくする。回路上では電圧をあげることに、アンプではボリュームを上げることに対応する。即ち $f(t)$ に対して、ある定数 $C > 1$ を用いて $\check{f}(t) = Cf(t)$ を対応させる作用である。

ふたつ目にクリッピングを行う。サチュレーションを発生させるとも言う。許容範囲を絞ることで、微分不可能な点や、矩形波に近い成分が含まれるようになる。即ち、 $\tilde{f}(t) = \max\{-1, \min\{1, \check{f}(t)\}\}$ を与える作用である。これにより微分不可能な点や高周波成分が大きくなり、歪んだ音となるのである。Figure 1 にディストーションエフェクタの各処理時における信号の概形を示す。

以下、信号にディストーションフィルタを作用させることを、ディストーションをかけるといい、ディストーションをかけた音を、ディストーションサウンドと呼ぶ。

1.2 ディストーションサウンドの解析

ギターサウンド等に用いられるディストーションエフェクタは、クリッピングを行うため、非線形・非可逆変換である。このフィルタにより生成される波形の倍音周波数のうち、原音の周波数レスポンスを H_1 、 N 次の倍音のレスポンスを H_N とする。歪みの度合いの指標として、全高調波歪み率 (Total Harmonic Distortion, THD) というものがある [2], [3]。これは

$$D_{\text{THD}} = \frac{\sqrt{H_2^2 + H_3^2 + \cdots + H_N^2}}{\sqrt{H_1^2 + H_2^2 + H_3^2 + \cdots + H_N^2}}$$

または

$$D_{\text{THD}} = \frac{\sqrt{H_2^2 + H_3^2 + \cdots + H_N^2}}{H_1}$$

と定義され、%もしくは比率として dB で表される。しかし、この定義によると、元の楽器音に多くの倍音成分が含まれていた場合の考慮がされておらず、幅広い音色に対してのディストーションサウンドを考えるには不適切である。

2 ディストーションサウンドの特徴量について

我々はウェーブレットを用いて、ディストーションサウンドの特徴量抽出を行っている。特に本研究では局在性の高い Haar ウェーブレットを用いて、ディストーションサウンドの特徴量を定義している [6],[7]。以下、信号 f は音高が固定されたもので、 $f \in L^1(\mathbf{R}) \cap L^2(\mathbf{R})$ であり、コンパクトサポートを持つとする。また、 ψ は Haar ウェーブレット

$$\psi(t) = \begin{cases} 1 & (0 < t \leq \frac{1}{2}), \\ -1 & (\frac{1}{2} < t \leq 1), \\ 0 & (\text{otherwise}) \end{cases}$$

を表すものとする。

2.1 微分法

1 つ目は波形の傾きに着目した特徴量である。クリッピングされた部分には元の音の情報が残っていないと考え、振幅の増加を特徴量として抽出する。信号 f に対して、具体的な特徴量 E_1 を

$$E_1(a) = \max_b a^{-1} \left| \int_{\mathbf{R}} f(t) \psi \left(\frac{t-b}{a} \right) dt \right| \quad (1)$$

と定義し、この値が大きいほど歪んでいる、即ちディストーションがかかっていると判断する。この特徴量を微分法と呼ぶことにする。与えられた信号 f に対して、Haar ウェーブレットとの相関をとることで、微分を積分変換として捉えることができる。(1) 式において a^{-1} を掛けないもの考えると、Haar ウェーブレットのスケールが変化すると、それに合わせて特徴量の値が大きく変化してしまう。そのため、積分後にスケールパラメータ a^{-1} を掛けることにより、Haar ウェーブレットのスケールの変化による特徴量の大きな変化を抑えることができる。これは、 $f(ax+b)$ の点 b における 1 次近似 $f(ax+b) \approx f(b) + af'(b)x$ を考えたときに、 f' の情報を引き出している、ということである。また、 a^{-1} は $\psi((t-b)/a)$ の $L^1(\mathbf{R})$ における正規化をする際に現れる係数と見ることもできる。

この特徴量の定義は、音の定常、非定常によらない、ということである。即ち、音が段々小さくなっていくような状況であっても、定常状態のものと同じように解析ができるのである。一方、この特徴量 E_1 にはクリッピングの情報は一切含まれていない。しかし、非常にシンプルな計算で済むので、有用性は高いと考えられる。

2.2 積分法

2 つ目は波形における面積に着目する。 L^2 ノルムをディストーションサウンドの尺度にする手法である。

$$E_2 = \int_{\mathbf{R}} |f(t)|^2 dt \quad (2)$$

と定義し、この値が大きいほど歪んでいるとする。この特徴量を積分法と呼ぶことにする。ディストーションの増幅する操作により、クリッピングされていても元の音と比較して面積が増えるので、それを特徴量として定義する。 $-1 \leq f(t) \leq 1$ であることから、信号を2乗することにより、クリッピング部分がよりはっきりと現れるため、ディストーションがかかっているものとかかかっていないものの差が明瞭に現れる。そのため、1つの固定された信号に対して、様々なレベルのディストーションがかかったものが存在した場合、この特徴量を用いたディストーションサウンドの解析は効果的である。一方、原音に倍音成分が多く含まれる楽器（ギター・金管楽器等）は、音を出す度に波形が大きく異なる。このような楽器の音に対してのディストーションサウンドに対して、この特徴量 E_2 は実用的でない。また、定常状態でない音に関しては、1周期分に対してのこの特徴量を考えて、比較する必要がある。

2.3 微分積分法

3つ目は、クリッピング部分に特徴量を見る方法である。

$$E_3(a) = \int_{\mathbf{R}} \left(\int_{\mathbf{R}} |f(t)|^{1/4} \psi \left(\frac{t-b}{a} \right) dt \right) db \quad (3)$$

と定義し、この値が小さいほど歪んでいるとする。これは $\int_{\mathbf{R}} \psi(t) dt = 0$ を利用したもので、クリッピングされた部分が定数関数に近くなることから、クリッピングされた部分が多いほど、この特徴量は小さくなるのである。微分法と積分法と比較して、明確にクリッピングの量によって差が現れるのがこの特徴量である。信号 f を $1/4$ 乗しているのは、クリッピング部分と、音の波形の変化部分の差をはっきりとさせるためである。この特徴量も非定常な信号に対してはあまり実用的でないため、1周期分に対して特徴量を考える必要がある。

3 微分法を用いたディストーションサウンドの音色近似

ここでは、微分法を用いたディストーションサウンドの解析方法を提案する。

3.1 Haar ウェーブレットのスケールと特徴量について

いくつかの歪みの度合いの異なるディストーションサウンドを準備し、適当なスケールの Haar ウェーブレットを用いて、スケールによってどの程度特徴量が増えるのかを観察した。

Figure 3 に、観察に用いた各音の波形を示す。一番上が純音の波形（sin 波）、上から二番目がエレキギターのエフェクタをかけていないもの、そこから順にディストーションエフェクタが強くなっていくように並べてあり、一番下が矩形波である。この6つの波形に対して、提案手法の中で、微分法における特徴量を抽出したのが Figure 4 である。

Figure 3 の上から6つ波形が、Figure 4 の横軸の左から6つの値それぞれと順に対応している。一番上のグラフが、 ψ の台が波長の $1/10$ となるようにスケール $a > 0$ を選んで、各波形の特徴量を抽出したものである。そして、上から2番目のグラフは1番上のグラフのウェーブレット ψ のスケールをさらに半分にして特徴量を抽出したグラフ。3番目は2番目のグラフから、更にスケールを半分にしたもの。4番目は3番目のグラフから更にスケールが半分のものを用いて、それぞれ特徴量を抽出したものである。また、一番下は各音の全高調波歪み率についてのグラフである。

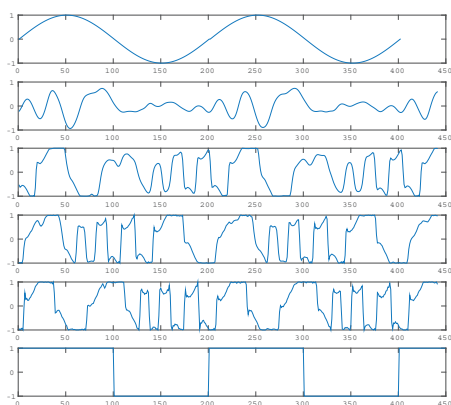


Figure 3: 観察に用いた音の波形

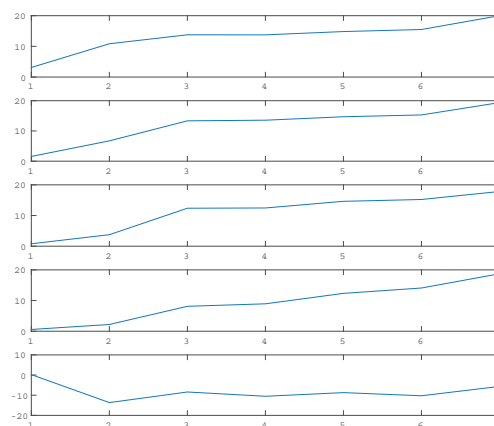


Figure 4: 各波形の特徴量

一番下を除く，提案手法を用いた各グラフを見てみると，全て右肩上がりになっているので，ディストーションサウンドの歪みの特徴量を抽出できていることがわかる．また，Haar ウェーブレット ψ のスケールに依らずに右肩上がりになっていることから，Haar ウェーブレットのスケールを適切に選べば，それよりも小さいスケールにの Haar ウェーブレットを用いても特徴量の差が得られることがわかる．

また，グラフをさらに観察すると，スケールが小さい Haar ウェーブレットを用いたほうが，よりディストーションのかかっている音の特徴量の傾きが大きくなっている．即ち，ディストーションのより強くかかっている音同士（グラフの右側）の差異をよりよく抽出していることがわかる．これは，ディストーションが大きくかかっている音には，より大きな増幅がかかっているため，Haar ウェーブレットにより，波形のクリッピングされていない，傾きの大きな部分の特徴量として抽出してくれているためである．

一方，Figure 4 の一番下，全高調波歪み率を見てみると，ディストーションサウンドの歪みを反映できていないことがわかる．すなわち，ディストーションサウンドの歪みの判定について，全高調波歪み率は利便性が低いことがわかる．

これらのことから，提案手法によるディストーションサウンドの特徴量を用いることにより，2つのディストーションサウンドの差異を判断する際に，次の手法を考えることができる：

手順 1. (2) 式，またはディストーションサウンドの音高に合わせて適当なスケールの Haar ウェーブレットを用いて，(1)，(3) 式により特徴量を抽出し，ディストーションのかかっている度合いを判断する．

手順 2. (1)，(3) 式において，2つのディストーションサウンドの特徴量が大きく異ならないとき，Haar ウェーブレットのスケールを縮めることで，ディストーションのかかり具合をより精密に判断する．

手順 3. ディストーションサウンドの特徴量に差ができるまで，手順 2 を繰り返す．

ディストーションが強くかかっている音の特徴量を抽出する際には，スケールが小さい

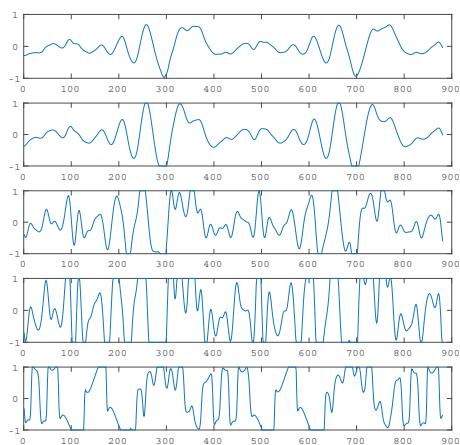
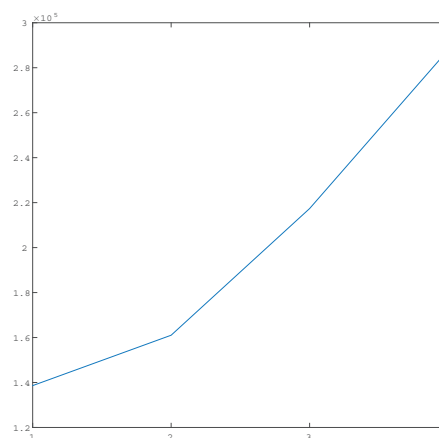


Figure 5: 近似実験結果

Figure 6: 近似実験の波形のスペクトルの相互
相関比較

Haar ウェーブレットを用いることで、より明確に差がでる。故に、より歪んでいる2つの音に対して、その比較をする際には手順2の回数が増えることになる。

3.2 ディストーションサウンドの音色近似手法

上で解説した、ディストーションサウンドの特徴量の比較手法を、実際の音色近似に用いることを考える。手順は以下の通りである。

手順 A. 音色近似をしたいディストーションサウンドの音高に合わせて、Haar ウェーブレットのスケールを適当に選び、(1), (2), (3) 式を用いてディストーションの特徴量を抽出する。

手順 B. スケールを変え、手順 A と同様にして音色を近似したいディストーションサウンドの特徴量を抽出する。

手順 C. 手順 B を適当な回数繰り返す。

手順 D. 得られた特徴量と一致するように、手元のエフェクタのつまみを調整し、音色を近似する。

3.3 近似実験とその結果

上の手順 A～手順 D を、微分法に対して用いて音の近似を行った。その結果を Figure 5 に示す。一番上が元となる音の波形であり、増幅とクリッピング、及びイコライジングにより音質を変化させ、特徴量を1つ合わせたものが上から2番目のグラフ、特徴量を2つ合わせたものが3番目のグラフ、3つ合わせたものが4番目のグラフ、そして1番下のグラフが近似したい目標の音の波形である。

Figure 5 を見ると特徴量を合わせるほどに、波形が近づいて行く様子が伺える。近似した波形のスペクトルと、近似したい波形のスペクトルの相互相関の最大値を比較したものが

Figure 6 である。Figure 5 の上から 1, 2, 3, 4 番目の波形が Figure 6 の横軸の各値 1, 2, 3, 4 と対応しており，縦軸がそれぞれの波形と，近似したい波形との相互相関の最大値を表している。これを見ると，右肩上がりになっていることが見て取れる。即ち，手順 C を繰り返し用いて，特徴量を複数個合わせることで，音質をよりよく近似できるということがわかった。

4 まとめ

Haar ウェーブレット， L^2 ノルムを用いて，ディストーションサウンドの特徴量を定義した。また，提案手法におけるディストーションサウンドの特徴量を，上で述べた手順 1 ～手順 3 で用いることで，ディストーションエフェクタの歪みを比較することができる。さらに，手順 A ～手順 D をふむことで，ディストーションサウンドの音色近似をすることができる。

References

- [1] Temme, Steve. "Audio distortion measurements." Application Note, Bruel, Kjar (1992).
- [2] Cabot, Richard C. "Fundamentals of Modern Audio Measurement," J. Audio Eng. Soc., Vol. 47, No. 9, Sep., 1999, pp. 738-762, Audio Engineering Society, NY, 1999.
- [3] D. Shmilovitz, "On the Definition of Total Harmonic Distortion and Its Effect on Measurement Interpretation", IEEE Transactions on Power Delivery, VOL. 20, NO. 1, January 2005.
- [4] I. Daubechies, "Ten lectures on wavelets", CBMS-NSF Regional Conference Series in Applied Mathematics, 61, SIAM, Philadelphia, PA, 1992.
- [5] G. Bachmann, L. Narici , E. Beckenstein, "Fourier and Wavelet Analysis", Springer Science+Business Media New York, 2000.
- [6] 鈴木 俊夫, 善甫 啓一, 木下 保, "ウェーブレット変換に基づくディストーションサウンドの特徴量抽出", 音響音楽研究会資料, Vol. 35 No.2 MA2016-09, 2016.
- [7] 鈴木 俊夫, 善甫 啓一, 木下 保, "ディストーションサウンドの Haar ウェーブレットに基づく音色近似", 音響音楽研究会資料, Vol. 35 No.4 MA2016-23, 2016.

DCT とウェーブレット変換を縦横に組み合わせた 直交変換をブロック毎に用いる 適応的画像圧縮法

芦澤 恵太^{*} 森田 雅貴[†]

^{*} 舞鶴工業高等専門学校 [†] 名城大学大学院都市情報学研究科

概要. 現在, 普及している非可逆画像圧縮方式の多くは, ブロック単位の離散コサイン変換 DCT に基づいている. これまでに我々は, 最も単純なウェーブレット変換であるハール変換を利用することで, DCT に由来する JPEG 標準の代表的な視覚的な歪の軽減を試みてきた. 本講演では, 我々の最近の研究成果も含め, ブロック単位での処理という枠組みの中で, ウェーブレット変換を利用する問題点と利点を整理し, 研究の発展可能性について議論する.

DCT and DWT based Block Adaptive Image Compression Algorithm Using the Quantization Errors in Each Block

Keita Ashizawa^{*} Masaki Morita[†]

^{*}National Institute of Technology, Maizuru College [†]Meijo University

Abstract. Currently, the most popular lossy image compression scheme is the JPEG, which is based on the DCT. A key issue in the JPEG is how to avoid edge artifacts due to the Gibbs phenomenon. To avoid the problem, we apply a frequency transform scheme partially based on the Haar transform to the JPEG. The Haar transform is particularly effective for the reduction of edge artifacts owing to its rectangular basis functions. Numerical examples using 8 standard images with various features show that our proposal improves the performance of the JPEG, especially for step edges.

1. はじめに

一般家庭で目にする画像の画素数の急激な増加は, 4K 規格のディスプレイやビデオカメラの急速な普及に顕著に現れている. 画素数の増加は, メモリ使用量や計算時間といった計算資源に対して大きな負荷となる為, 高効率で高速な非可逆画像圧縮方式の開発は急務である. 圧縮画像において視覚的に精細感を得る為には, グラデーション部分のなめらかな表現とオブジェクトと背景との境界 (エッジ) のような線的な特徴をはっきりと表現することの両立が欠かせない.

既存のデジタル画像圧縮の多くは, ブロック単位の離散コサイン変換 (DCT) に基づく方式が広く用いられる. しかし, DCT を用いた場合, 画像を再構成する為の基底がブ

ブロック全体を覆う為、エッジを含むブロックでは、歪がブロック全体に拡がり視覚的な劣化を引き起こすことが知られている。そこで我々は、基底が局在する離散ウェーブレット変換 (DWT) に着目した。これまでに、矩形波を基底とし最も単純な DWT であるハール変換 (HT) を DCT に組み合わせた新たな基底系の画像圧縮への応用について研究してきた。本稿では、前回の本ワークショップ以降に提案してきた基底系の特徴についてまとめると共に、それら基底系を画像圧縮に適応する際の問題点を整理する。その上で、ブロック単位での処理という枠組みの中で、ウェーブレット変換をどのように活用していくのかということを経験したい。

2. 組み合わせ式直交変換 (COT)

これまで、異なる直交変換を同一ブロックで組み合わせることは、対称性を欠くため避けられてきた。我々は、直交変換を組み合わせた複数の基底系を予め辞書として用意し、その中で最適なものをブロック毎に選択することで失われた対称性を補うことを試みている [1,2]。以下では、この方式を組み合わせ式直交変換 (COT: Combinational Orthogonal Transform) 方式と呼ぶ。

2.1 概略

入力画像を輝度信号とした場合の直交変換に基づく圧縮画像作成の概略を Fig. 1 に示す。JPEG 標準方式では、 8×8 画素単位でブロックに分割されたすべてのブロックにおいて、DCT を適用し得た変換係数を量子化テーブルを用いて量子化した後に符号化を行う [3]。それに対して、COT 方式では、ブロック毎に適応的に直交変換を行う。Fig. 2 に COT 方式の概略を示す。Fig. 2 における基底系番号 k とは、各ブロックで行う直交変換と対応しており、量子化された変換係数と合わせて伝送される。すなわち、符号時は直交変換を組み合わせ構成した基底系それぞれの変換係数を算出する必要があるが、復号時は基底系番号 k により指定された逆変換のみを行う。

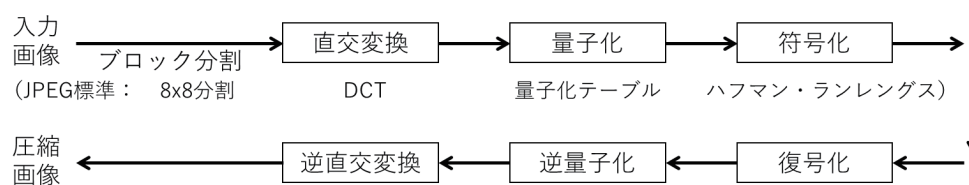


Fig. 1. 従来方式

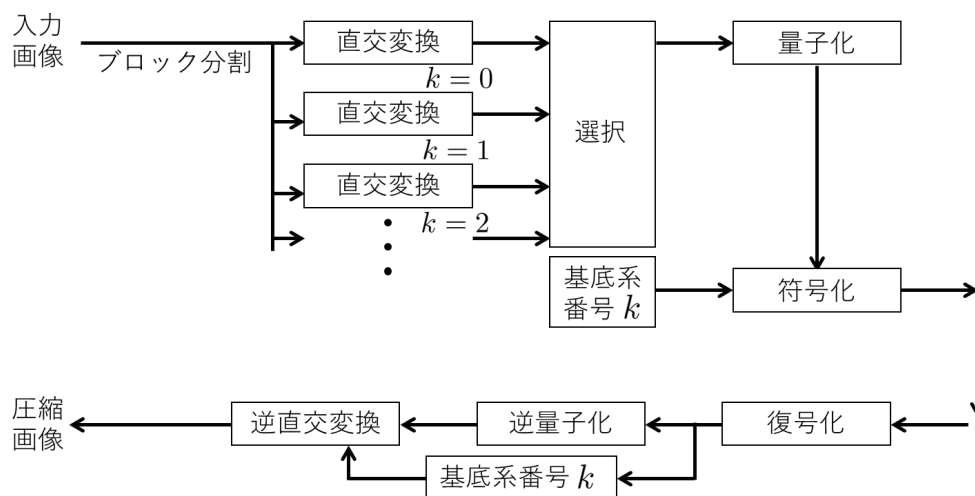


Fig. 2. COT に基づく適応的な画像圧縮方式

2.2 DCT と HT を組み合わせた基底系

文献 [1] では, JPEG 標準等で用いられている 2 次元 DCT を初期直交変換とし, Fig. 3 に示す 3 種類の基底系をブロック毎の選択肢として与えた. Fig. 3 の **DCT-DCT** は 2 次元 DCT の基底系 $k=0$ を表しており, **DCT-HT** および **HT-DCT** は, それぞれ, **DCT-DCT** において, 水平方向のみを 1 次元 HT に, 垂直方向のみを 1 次元 HT に置き換えて構成した基底系である. さらに **NHT** は非標準型 HT の基底系を表している.

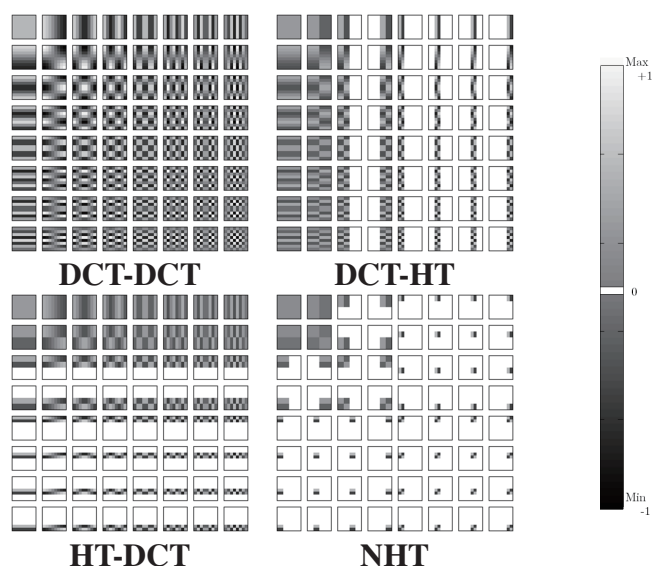


Fig. 3. 2 種類の直交変換 (DCT, HT) を組み合わせた基底系

2.3 DCT と HT とアルパート変換を組み合わせた基底系

静止画像，すなわち，2次元信号を対象とした場合，垂直と水平方向にどのような直交変換を採用するのかを考えた場合，COT方式では2種類以上の直交変換を組み合わせることが可能となる．文献[4]では，アルパートウェーブレット変換(AT)[5,6]を含めた場合の構成例を報告している．3種類の直交変換の内2種類を縦横に組み合わせることで，基底系の選択肢は9種類に増加する．Slant-Haar変換[7]とも呼ばれる低次のATを加えた場合は，Fig. 3の4種類に，Fig. 4に示す5種類の基底系が選択肢に加わる．

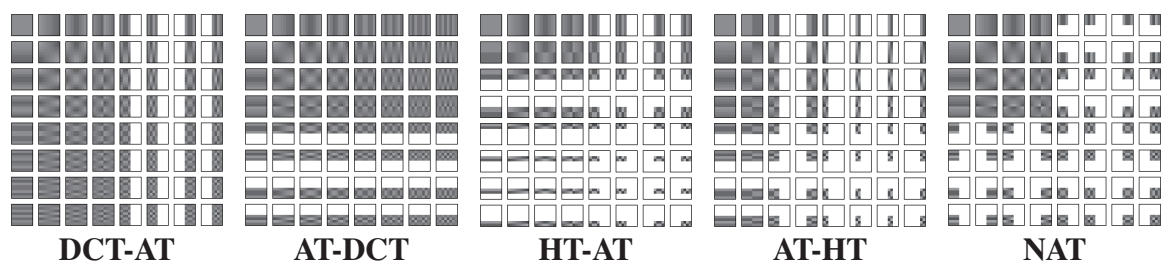


Fig. 4. アルパート変換(AT)を組み合わせた基底系

COT方式では，量子化された変換係数と合わせて，各ブロックでどの基底系に対応した変換が選択されたのかという情報のみを送れば良い．水平垂直方向の各処理において， N 種類の直交変換を候補とした場合，組み合わせで生じる基底系 N^2 種類が選択肢の最大となる．この場合には， N^2 種類の基底系の中から画像に応じて選択肢を絞ることで，付加情報量を抑制することが出来る．例えば，1回目の選択で頻出した上位4種類の基底系のみで再選択を行うことで，基底番号 k の情報量は，可逆符号化を行う前であっても，各ブロックあたり2bitとなる．文献[8,9]では，ブロックごとに水平方向と垂直方向を代表するベクトルを算出し，基底系を構成していく方式が提案されているが，この場合は代表ベクトルも付加情報となる．COT方式の最大の特徴は，既存の(有名な)1次元変換の組み合わせで選択肢を構成することで，ブロック単位での適応的な処理ではあるものの，付加情報量を抑制する点である．

3. COTに基づく画像圧縮方式の実装

文献[1,2,4]では，水平垂直に異なる直交変換を組み合わせた新たな基底系の画像圧縮への有効性を検証する為に，暫定的な方法によって基底系の選択および一様量子化を行ってきた．また，情報量の算出においても，符号化は行わず平均情報エントロピーを採用してきた．これまでの研究により，いくつかの新たな基底系を構成すると共に，COT方式の枠組みを確定することができた．そこで本節では，Fig.3に示すDCTとHTを組み合わせた基底系に基づくCOT方式に対し，具体的な選択・量子化の構成例を報告する．

3.1 COT 方式 (DCT, HT) における量子化テーブルの構成例

入力画像を 8×8 画素ブロックに分割し、分割されたそれぞれのブロックの画素値からなる 8×8 行列を f_ℓ ($\ell = 1, 2, \dots, L$) とする. ここで, L は入力画像全体に含まれるブロック数, ℓ はそれぞれのブロックを識別するための番号である. Fig. 3 の基底系 **DCT-DCT** を $k = 0$, 基底系 **DCT-HT** を $k = 1$, **HT-DCT** を $k = 2$, **NHT** を $k = 3$ とした場合, 各ブロックにおける変換係数を $F_\ell^{(k)}$ と表す. ここでは, k 番目の変換における量子化定数の重み付けを行列 $W^{(k)}$ を用いて表す. まず $k = 0$ の場合は 2 次元 DCT 係数の量子化なので, $W^{(0)}$ には標準的な JPEG 圧縮の量子化テーブルを用いるものとする [3]. 一方で, Fig.3 に示すようにハール変換の基底系は階層ごとに共通の形状からなる基底によって構成されるため, $k = 1, 2, 3$ の場合に JPEG 圧縮の量子化テーブルをそのまま用いるのは適切ではない. そこでハール変換を含む場合には, 階層ごとに行列 $W^{(0)}$ の算術平均を用いることによって対応する. 量子化定数の重み付けに用いる行列 $W^{(0)}$, $W^{(1)}$, $W^{(2)}$, $W^{(3)}$ を以下に示す.

$$W^{(0)} = \begin{pmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \end{pmatrix}, \quad W^{(1)} = \begin{pmatrix} 16 & 11 & 13 & 13 & 44 & 44 & 44 & 44 \\ 12 & 12 & 17 & 17 & 50 & 50 & 50 & 50 \\ 14 & 13 & 20 & 20 & 56 & 56 & 56 & 56 \\ 14 & 17 & 26 & 26 & 70 & 70 & 70 & 70 \\ 18 & 22 & 47 & 47 & 89 & 89 & 89 & 89 \\ 24 & 35 & 60 & 60 & 98 & 98 & 98 & 98 \\ 49 & 64 & 83 & 83 & 111 & 111 & 111 & 111 \\ 72 & 92 & 97 & 97 & 104 & 104 & 104 & 104 \end{pmatrix},$$

$$W^{(2)} = \begin{pmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 15 & 19 & 27 & 46 & 72 & 75 & 59 \\ 14 & 15 & 19 & 27 & 46 & 72 & 75 & 59 \\ 41 & 53 & 66 & 76 & 91 & 109 & 110 & 92 \\ 41 & 53 & 66 & 76 & 91 & 109 & 110 & 92 \\ 41 & 53 & 66 & 76 & 91 & 109 & 110 & 92 \\ 41 & 53 & 66 & 76 & 91 & 109 & 110 & 92 \end{pmatrix}, \quad W^{(3)} = \begin{pmatrix} 16 & 11 & 15 & 15 & 55 & 55 & 55 & 55 \\ 12 & 12 & 15 & 15 & 55 & 55 & 55 & 55 \\ 15 & 15 & 23 & 23 & 55 & 55 & 55 & 55 \\ 15 & 15 & 23 & 23 & 55 & 55 & 55 & 55 \\ 59 & 59 & 59 & 59 & 100 & 100 & 100 & 100 \\ 59 & 59 & 59 & 59 & 100 & 100 & 100 & 100 \\ 59 & 59 & 59 & 59 & 100 & 100 & 100 & 100 \\ 59 & 59 & 59 & 59 & 100 & 100 & 100 & 100 \end{pmatrix}.$$

JPEG に対応する市販のソフトでの画質設定を参考に, 圧縮画像の品質を制御するための係数 $q \in [0, 100]$ (以下, 画質設定値と呼ぶ) を導入する. 行列 $W^{(k)}(i, j)$ が i 行 j 列の位置にある要素を示すこととし, 量子化行列 $Q^{(k)}$ を以下で定義する.

$$Q^{(k)}(i, j) = \begin{cases} 255 & , \quad q \leq \frac{50}{255} W^{(k)}(i, j), \\ \text{rd}\left(\frac{50}{q} W^{(k)}(i, j)\right) & , \quad \frac{50}{255} W^{(k)}(i, j) < q \leq 50, \\ \text{rd}\left(\frac{100 - q}{50} W^{(k)}(i, j)\right) & , \quad 50 < q \leq 100 - \frac{50}{W^{(k)}(i, j)}, \\ 1 & , \quad q > 100 - \frac{50}{W^{(k)}(i, j)}. \end{cases}$$

ここで, $\text{rd}(\cdot)$ は四捨五入の操作を表す. すなわち, 画質設定値 q が与えられた場合, 各ブロックにおける量子化係数 $\tilde{F}_\ell^{(k)}(i, j)$ は, 選択された基底系 k , および, 対応する重み行列

$W^{(k)}$ に基づき算出される量子化行列 $Q_\ell^{(k)}(i, j)$ によって以下で与えられる.

$$\widetilde{F}_\ell^{(k)}(i, j) = \text{rd} \left(\frac{F_\ell^{(k)}(i, j)}{Q_\ell^{(k)}(i, j)} \right).$$

3.2 圧縮画像の誤差評価に基づく基底系選択の構成例

JPEG 標準では, 高周波成分に重み付けされた量子化により大半の高周波成分は捨てられ, 圧縮画像は主に低周波成分から再構成される. この点を考慮すると, 基底系の選択は, 圧縮画像の再構成に用いられる低周波成分が抽出された後, すなわち, 量子化後に実行する方が効果的であることが期待される. そこで, ブロック毎に各基底系に対する量子化誤差の 2 乗和 $d_\ell^{(k)}$:

$$d_\ell^{(k)} = \sum_i \sum_j \left(F_\ell^{(k)}(i, j) - Q_\ell^{(k)}(i, j) \widetilde{F}_\ell^{(k)}(i, j) \right)^2$$

を算出し, 値が最も小さくなる基底系を選択する. 各ブロック $\ell = 1, 2, \dots, L$ において選択された基底系番号を K_ℓ と表記すれば, 上述の手順は以下にまとめられる:

Initialize $K_\ell = 0$.

For $k = 1, 2, 3$, set $K_\ell \leftarrow k$ if $d^{(k)} < d^{(K_\ell)}$.

Fig.3 に示すように, 入力画像に含まれる全ブロックで算出された $\widetilde{F}_\ell^{(K_\ell)}$ にいて, JPEG 方式と同様のエントロピー符号化を適用する. 選択された基底系番号 K_ℓ については, 直流成分 $\widetilde{F}_\ell^{(K_\ell)}(0, 0)$ 以外の成分を持つ場合, 言い換えるならば, $\widetilde{F}_\ell^{(K_\ell)}$ の交流成分に非零の値を持つブロックの K_ℓ のみにハフマン符号化を適用する. これは, 係数行列 $\widetilde{F}_\ell^{(K_\ell)}$ の直流成分は, 明らかに K_ℓ に依らず同一の値を持つためである.

3.3 数値実験

3.1 節および 3.2 節の構成例 (以下, OUR と表記) の有効性を, Fig. 5 に示す 8 種類のテスト画像を用いた数値実験で確認する. 比較対象は, DCT のみを用いた場合 (DCT), および, 変換係数の 1-ノルムの値を最小とする基底系を選択した場合 (1-norm) [1] である. 圧縮画像の画質評価には, PSNR 値及び MSSIM 値を用いた. PSNR 値は, 入力画像と圧縮画像の平均 2 乗誤差 RMSE から次式で与えられる:

$$\text{PSNR} = 20 \log_{10}(255/\text{RMSE}).$$

MSSIM 値は, 画素値とコントラストに対して正規化された局所的パターンを比較する誤差評価指標であり, PSNR 値に比べてより見た目を重視する指標として知られている [10]. PSNR 値と MSSIM 値はともに値が大きいほど圧縮画像の画質が良いことを意味する. 画

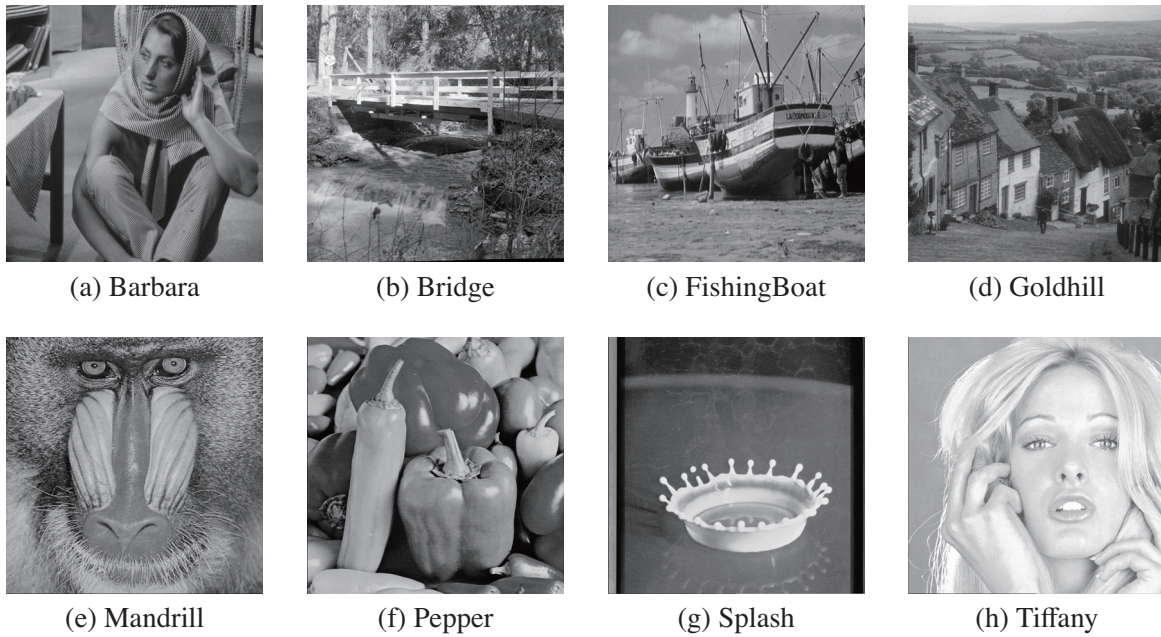


Fig. 5. テスト画像 (512 × 512 画素, 8 bits/pixel グレースケール)

画質設定値 q を 40, 50, 60 とした際のビットレートに対する PSNR 値をプロットしたものを Fig.6 に, MSSIM 値をプロットしたものを Fig.7 に示す. ここで, OUR と 1-norm のビットレートは, 非零値の交流成分が量子化後に残ったブロックの基底系選択番号 K_ℓ の符号を加えたレートである. 図 6, 図 7 より, 先行研究の 1-norm に比べ, 同一のビットレートに対する PSNR 値及び MSSIM 値については OUR の方が高いことが分かる.

次に, 各テスト画像の圧縮においてそれぞれのブロックに各基底系がどのように配分されたかを Tab. 1 に示す. Tab. 1 には, $q = 50$ で圧縮画像を再構成した際の各基底系の配分を示しており, $N_{OUR}^{(K_\ell)}$, $N_{1-norm}^{(K_\ell)}$ は, それぞれ OUR, あるいは 1-norm で $K_\ell = k$ の基底系が選択されたブロックの個数を表している. 表 1 では各基底系の配分が OUR と 1-norm で大きく変化しているが, Fig. 6, Fig. 7 の結果を考えると, 同一のビットレートに対する PSNR 値と MSSIM 値を高くするためには, OUR の選択方法による基底系の配分が有効であることが見て取れる.

従来の JPEG 方式に対応する DCT と比較し, エッジ近傍の歪が顕著に改善された例を Fig. 8 に示す. Fig. 8 は上から Barbara, Bridge, FishingBoat の順に, 入力画像, OUR による圧縮画像, DCT による圧縮画像から, 128 × 128 画素領域を切り出したものである. また, DCT による圧縮画像は画質設定値 $q = 50$ のときのものであり, OUR については DCT 圧縮画像全体に対するビットレートと小数点以下第 2 位まで揃うように q の値を定めている. それぞれのビットレートの値は図中のキャプションに示してある. Fig. 8 の (a)Barbara の DCT 圧縮画像では, 縞模様のストールに隣接する滑らかな顎の部分に入力画像にはない縞模様の歪が見られる. また, Fig. 8 の (b)Bridge の DCT 圧縮画像においては, 橋の側面の白い部分に入力画像にはない縞模様の歪が見てとれる. 更に, Fig. 8

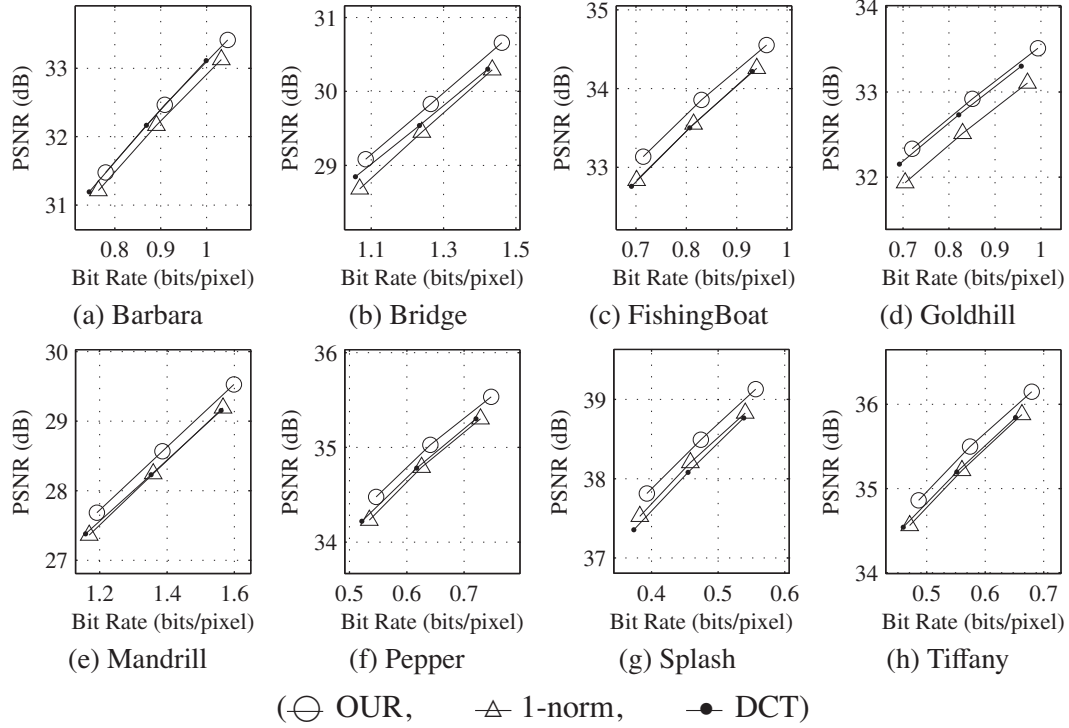


Fig. 6. PSNR 値を用いた圧縮性能の比較

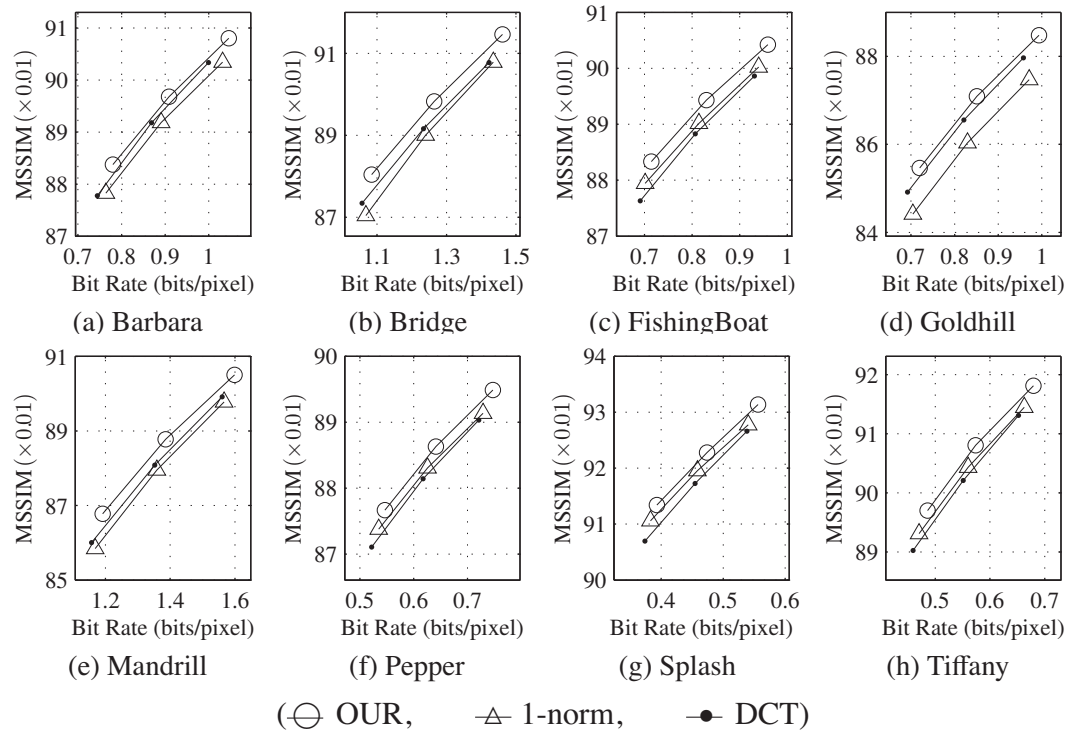


Fig. 7. MSSIM 値を用いた圧縮性能の比較

Table 1. 各基底系が割り当てられたブロック数 ($q = 50$)

		K_ℓ			
		$k = 0$	$k = 1$	$k = 2$	$k = 3$
(a) Barbara	$N_{OUR}^{(K_\ell)}$	2453	1060	327	163
	$N_{1-norm}^{(K_\ell)}$	2151	715	212	856
	$N_{OUR}^{(K_\ell)} - N_{1-norm}^{(K_\ell)}$	302	345	115	-693
(b) Bridge	$N_{OUR}^{(K_\ell)}$	2519	730	633	212
	$N_{1-norm}^{(K_\ell)}$	1702	760	971	658
	$N_{OUR}^{(K_\ell)} - N_{1-norm}^{(K_\ell)}$	817	-30	-338	-446
(c) FishingBoat	$N_{OUR}^{(K_\ell)}$	2380	1131	249	120
	$N_{1-norm}^{(K_\ell)}$	2028	1118	376	302
	$N_{OUR}^{(K_\ell)} - N_{1-norm}^{(K_\ell)}$	352	13	-127	-182
(d) Goldhill	$N_{OUR}^{(K_\ell)}$	2439	911	500	171
	$N_{1-norm}^{(K_\ell)}$	1546	818	666	970
	$N_{OUR}^{(K_\ell)} - N_{1-norm}^{(K_\ell)}$	893	93	-166	-799
(e) Mandrill	$N_{OUR}^{(K_\ell)}$	2343	525	1012	215
	$N_{1-norm}^{(K_\ell)}$	2004	503	1196	389
	$N_{OUR}^{(K_\ell)} - N_{1-norm}^{(K_\ell)}$	339	22	-184	-174
(f) Pepper	$N_{OUR}^{(K_\ell)}$	2361	655	737	211
	$N_{1-norm}^{(K_\ell)}$	2044	629	797	452
	$N_{OUR}^{(K_\ell)} - N_{1-norm}^{(K_\ell)}$	317	26	-60	-241
(g) Splash	$N_{OUR}^{(K_\ell)}$	1989	783	738	310
	$N_{1-norm}^{(K_\ell)}$	1596	827	772	555
	$N_{OUR}^{(K_\ell)} - N_{1-norm}^{(K_\ell)}$	393	-44	-34	-245
(h) Tiffany	$N_{OUR}^{(K_\ell)}$	2246	754	553	197
	$N_{1-norm}^{(K_\ell)}$	1971	802	512	385
	$N_{OUR}^{(K_\ell)} - N_{1-norm}^{(K_\ell)}$	275	-48	41	-188

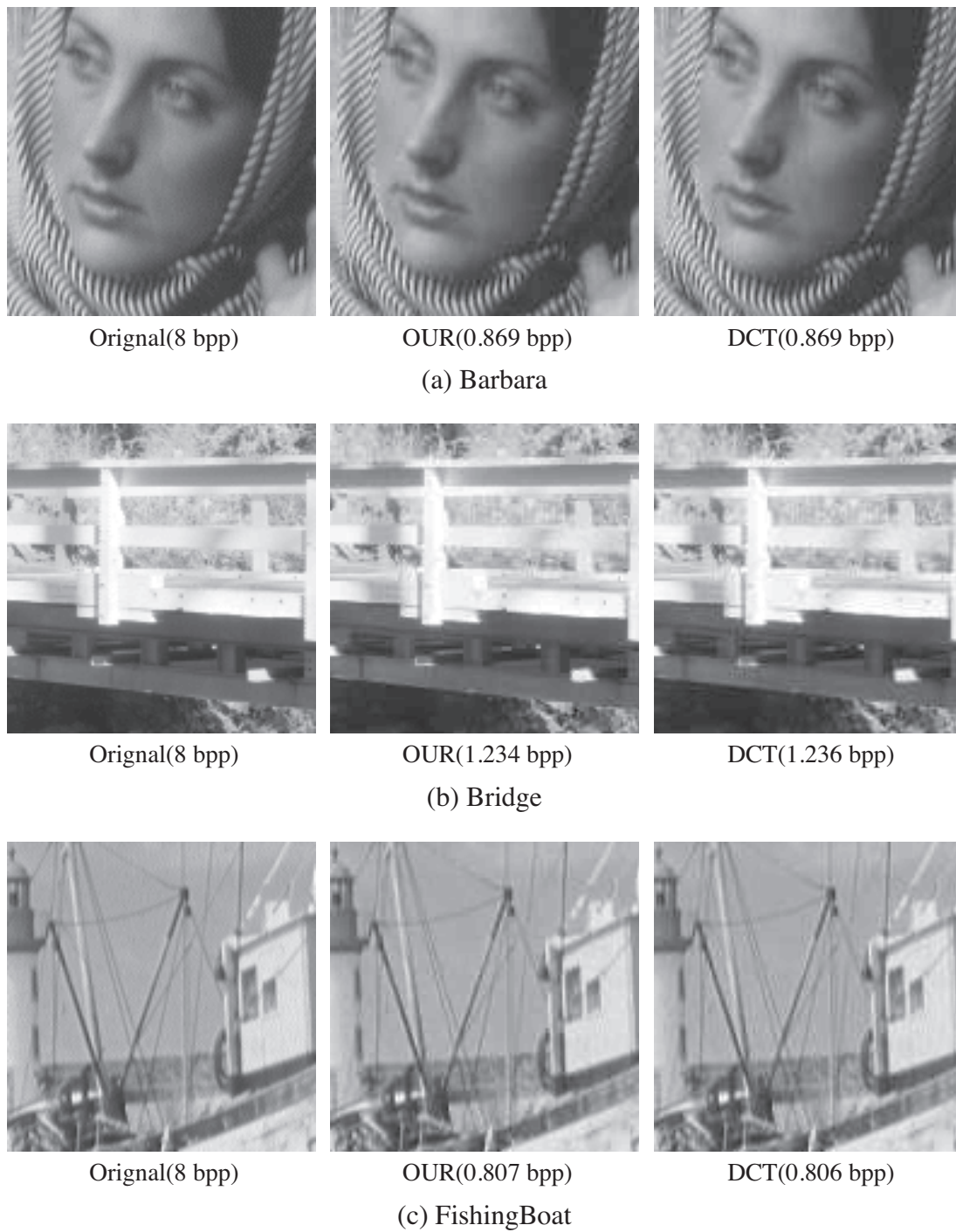


Fig. 8. OUR によって DCT 圧縮画像におけるエッジ近傍の歪が軽減された領域.

の (c)FishingBoat の DCT 圧縮画像においては、入力画像では滑らかだった背景の空に船のマスト及びロープの成分が映り込んでいる。DCT 圧縮ではブロック全体にフーリエコサイン級数近似を適用するため、ブロック画像内の一部の画素値が急激に変化すると、ブロック画像全体にギブス現象の影響が及ばされ、このような歪が生じてしまう。一方、提案方法による圧縮画像の同部位を見てみると、明らかにこれらの歪が軽減され、圧縮画像の画質が改善されたことが確認できる。

4. まとめと今後の課題

本稿では、我々がこれまでに提案してきた異なる直交変換を縦横に組み合わせて構成する基底系の画像圧縮への適用について考えてきた。2 節で述べた組み合わせ式直交変換 (COT) に基づく、ブロック単位で適応的に周波数変換を採用する方式は、その組み合わせの元となる直交変換を変えることで、様々な信号に対して適用することが可能となる。画像圧縮方式として実装するには、基底系の選択方法だけでなく、量子化や符号化といった各要素技術において、基盤となる構成を確立する必要がある。3 節で報告した構成例は、MSSIM 値を用いた圧縮画像の客観評価、および、主観評価において有効であることが確認できた。

我々の研究は、JPEG 方式の改善を目標としていた為、これまで DCT を基本に HT を組み合わせることで進めてきた。本研究を発展させていくために、画像圧縮以外の新たな応用分野の提案など、参加者の意見を聞きたい。

謝辞 本研究は、名城大学の山谷教授との共同研究である。また、本研究を進めるにあたり、プログラムの作成及び実験データの収集・整理に協力して頂いた山谷研究室の丹羽剛生氏（名城大学大学院都市情報学研究科）に、この場をお借りして心から感謝の意を表したい。

参考文献

- [1] 芦澤恵太, 山谷克, “DCT とハール変換を縦横に組み合わせた新たな周波数変換方式の提案と画像圧縮におけるモスキートノイズの軽減,” 信学論 (A), Vol.J96-A, No.7, pp.484-492, 2013.
- [2] 森田雅貴, 芦澤恵太, 山谷克: “DCT と HT に基づく組み合わせ式直交変換における演算量の軽減方法とその高精細画像圧縮への応用”, 都市情報学研究, 19, pp.81-92 , 2014.
- [3] W.B. Pennebaker and J.L. Mitchell, JPEG Still Image Data Compression Standard, Van Nostrand Reinhold, New York, 1993.
- [4] 芦澤恵太, 原田卓弥, 山谷克, “ブロック変換に基づく画像非可逆圧縮におけるウェーブレット変換,” 数理解析研究所講究録, 2001, pp.83-94, 2016.
- [5] B. Alpert : “A class of bases in L_2 for the sparse representation of integral Operators”, SIAM J. Math. Anal. 26, pp.246-262, 1993.
- [6] 芦野隆一, “マルチウェーブレットの構成と応用”, 数理解析研究所講究録, 1622, pp.18-36, 2009.
- [7] W.K.Pratt, W-H.Chen, L.R.Welch : “Slant Transform Image Codeing”, IEEE Trans. Commun., COM-22, pp.1075-1095, 1974.
- [8] S. Minasyan et al., “An image compression scheme based on parametric Haar-like transform,” Proc. ISCAS, vol. 3, pp. 2088-2091, 2005.
- [9] S. Minasyan, J. Astola, D. Guevorkian : “ On a Class of Parametric Transforms and Its Application to Image Compression”, Advances in Signal Processing, Article ID 58416, 14 pages, 2007.
- [10] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, “Image quality assessment, From error measurement to structural similarity,” IEEE Trans. Image Process., vol.13, no.4, pp.600-612, 2004.

芦澤 恵太 (舞鶴工業高等専門学校電気情報工学科)

〒625-8511 京都府舞鶴市字白屋 2 3 4 番地

E-mail: ashizawa@maizuru-ct.ac.jp

他覚的聴力検査に用いる誘発脳波の ウェーブレット解析についての考察

井川 信子^{*} 守本 晃[†] 芦野 隆一[†]

^{*} 流通経済大学 [†] 大阪教育大学

概要. 自分の意思で聞こえを応えることができない場合、他覚的聴力検査が実施される。その中でも耳から音刺激を与えて誘発される脳波を用いる検査がある。この検査はMRIなどに比べると安価で行われる検査なので健診などでの活用が望まれるが、脳波解析の難しさのために検診レベルの実用化が遅れている。計測時即時診断の実現を求める目的で、我々はこれまで聴性誘発脳波解析にウェーブレット解析を適用してきた。これまでの成果と課題についてまとめることで、問題点を明確にし解決策について議論する。

Application of wavelet analysis to auditory evoked brain responses using objective audiometry test

Nobuko Ikawa^{*} Akira Morimoto[†] Ryuichi Ashino[†]

^{*}Ryutsu Keizai University [†]Osaka Kyoiku University

Abstract. Wavelet analysis is applied to the auditory filter of the cochlear models. Auditory evoked brain responses which obtained at mid brain that relay part of the auditory central system are used to assist human objective audiometry test. As examples are Auditory Brainstem Response (ABR) and Auditory Steady-State Response (ASSR). In our previous study we proposed a novel quick diagnosis by applying the discrete stationary wavelet analysis (SWT) to the waveforms of each averaging. In this study, we discuss the effectiveness and possibility of the wavelet analysis to accuracy and quickly detection of the auditory evoked responses.

1. はじめに

ウェーブレット解析ではある時間帯のみで振動し、それ以外ではゼロとなる波形を用いることで、周波数構造の時間変化を捉えにくいフーリエ解析に対して、時刻と周波数を同時に扱いたい場合、例えば時系列データ解析など工学における応用として広く活用されている(例えば [1])。聴覚特に末梢系である蝸牛基底膜振動系は、時間的にも周波数的にも非対称な応答特性をもつ帯域通過フィルタとみなされ、Patterson の Gamma Tone モデルや入野の Gamma Chirp モデルいわゆる聴覚フィルタモデル(例えば [2] 参照)は良い近似であることが知られている。つまり、フーリエ分析器というよりもウェーブレット分析器とみなされている。

一方、被検者自身がきこえるかきこえないかについて正確な意思表示ができない場合など、被検者の応答に頼らずに聴力の測定をする必要がある場合に実施される聴覚検査を他覚的聴力検査 (objective audiometry test) という [3]. 他覚的聴力検査のなかに、音刺激に対する脳波上の変化、特に蝸牛からさらに進んだ聴覚中枢系の反応である聴性誘発反応 (auditory evoked responses) を指標として聴力を測定する方法がある. 我々はそのなかでも特に、聴性脳幹反応 (Auditory Brainstem Response: 単に ABR という) や聴性定常反応 (Auditory Steady-State Response : 単に ASSR という) を中心にウェーブレット解析を用いて、より迅速な反応抽出やモデルの改善などを試みてきている. これらの反応は末梢系と高次聴覚脳系の間によるものであり、その機能等現在も不明な点が多い反面、他覚的聴覚検査等において広く用いられているものである.

本稿では、まず、ABR と ASSR について述べる. 次に、1 次元複素連続ウェーブレット解析 (One-Dimensional Complex Continuous Wavelet Analysis : 以下 “CCWA” と略す) の適用例、1 次元離散定常ウェーブレット解析 (One-Dimensional Discrete Stationary Wavelet Transform : 以下 “SWT” と略す) の適用例をそれぞれ示し、聴覚中枢系におけるこれらの聴性誘発反応に実施してきたウェーブレット解析の有効性と可能性について問題を提起してその解決策について議論する.

2. ABR と ASSR について

ABR(詳細は [4]) は蝸牛神経と脳幹部聴覚路由来の反応で音刺激を与えてから 10 ミリ秒程度以内に認められる. 中間 (潜時) 反応 (Middle Latency Response : MLR) は内側聴覚状態レベルから聴皮質由来の反応とされており、音を与えてから 100 ミリ秒程度以内に認められる. 脳波の反応成分にはそれぞれ名前がつけられていて、潜時 (音刺激を入力後反応が得られるまでの時間) が 50, 100, 200 msec である反応は、それらの潜時をとって、P50, N100, P200 などと呼ばれる. また、潜時の早い方から順番に P1, N1, P2 などとも呼ばれる. N と P は極性を示していて、それぞれ、陰性 (Negative), 陽性 (Positive) を意味している. すなわち、各波形のピークには、ポジティブの P とその潜時の順に番号が、各波形の谷には、ネガティブの N とその潜時の順に番号がつけられている. ABR は、第 I 波から第 VII 波の 7 個のピークを持つ波から構成され神経線維のインパルス放電 (刺激のない状態から刺激のある状態に、または一定の刺激状態から異なる刺激状態に変化したときにのみ生じる反応で、On-反応ともいう) であり、蝸牛神経核から下丘におよぶ脳幹の広い範囲の聴覚伝導路がその発生起源とみられ、ABR の第 I 波から第 V 波の起源と脳幹の聴覚伝導路、蝸牛→蝸牛神経→蝸牛神経核→上オリーブ核→外側毛帯核→下丘→内側膝状体→聴放線→聴皮質 の対応がほぼ同定されている.

Fig. 1(引用 [5]) の左図の上のグラフは音刺激間隔 (ISI) 600 msec, 音刺激音圧 90 dB のクリック音の反応波形である. 左下図の波形は、上と同じ波形の時間軸を対数尺度で表示したものである. 左上図で P0 は ABR (正確には slow ABR), 次のピークから順に、

$P1=Pa$, $P2=Pb$ である．左下図で示すように 10 msec までの部分に ABR の速波成分，100 msec の部分に MLR，500 msec の部分に SVR が認識できる．右図は聴覚伝道路上の ABR の各波とその起源の対応，MLR の起源の推定を表している [5]．

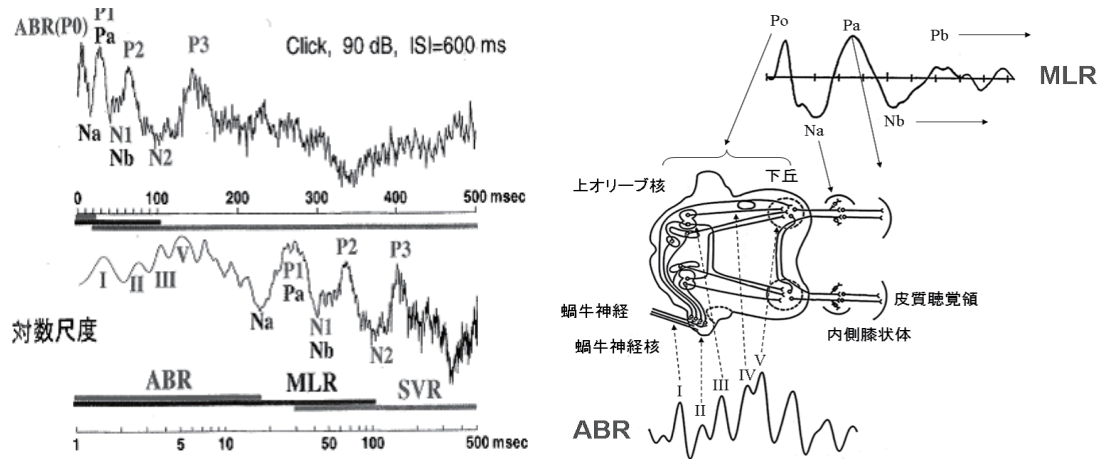


Fig. 1. Pathway of ABR and MLR [5].

2.1 ABR の計測

音刺激に誘発された脳波電位を時間的にサンプリングしてディジタル値に変換した数値を波形信号として用いる．音刺激ごとに得られた波形信号（これを 1 epoch という）のサンプリング時間を 10 ミリ秒，サンプリング点を 512 点，従って，サンプリング間隔は 0.019 ミリ秒（10 ミリ秒/512 点），これを繰り返し加算して，ABR 波形を得る．各加算回数における波形のディジタル値に変換したサンプリング値をファイルに出力した．

聴力健常の 19-21 歳男性 30 人から得られた実験データに基づいて解析を実施した．聴力正常者の場合でも，2000 回加算を繰り返せば，ABR のピークが必ず得られるというわけではないが我々の実験において 2000 回程度の加算すると，一般的に最も大きい振幅をもつ第 V 波の出現が得られることを手掛かりしている．

定義 2.1 Epoch_k は k 番目の epoch であり，

$$\text{ABR}_N = \frac{1}{N} \sum_{k=1}^N \text{Epoch}_k.$$

ABR_N を N -average ABR という．ここで，2000 - average ABR を単に ABR と呼ぶ．

ABR ではピーク潜時（刺激を与えてから誘発電位がピークに達するまでの時間）などの指標が臨床的に応用される．例えば音刺激音圧（音刺激強度，Intensity）を変更し第 V 波のピークの有無により閾値を求めて，推定オーディオグラムを描くことで難聴の診断を行

う、また、第 V 波のピーク潜時 (peak latency) を調べ、刺激音圧と第 V 波ピーク潜時の関係を表した曲線 (I-L curve) を描くことで、難聴の特徴を観察する。

2.2 ASSR の計測

ASSR とは 1 秒間に 40 回から 100 回の繰り返した聴覚刺激に対し脳波が定常的な反応をする状態である。ASSR には 40-Hz ASSR と 80-Hz ASSR があり、40-Hz ASSR の起源は slow ABR (P0) + MLR すなわち P0+Pa+Pb ([6] 参照)、80-Hz ASSR の起源は slow ABR すなわち P0 と考えられている。40-Hz ASSR は覚醒時に、80-Hz ASSR は睡眠時に観察される [5]。ABR に比べて周波数特異性が高い。睡眠時に観察される 80-Hz ASSR を用いて推定オーディオグラムを描くことなどで難聴の臨床診断等を行っている。

80-Hz ASSR の場合

80-Hz ASSR については、蝸牛聴覚フィルタモデル^{*1}を用いて、聴力検査を実施する方法がある。実際、睡眠時に 80-Hz ASSR の振幅が比較的大きいことを利用して Picton ら [7] により開発された聴力検査装置ナビゲータプロ & MASTER (Multiple Auditory Stady-State Response 以降、単に MASTER という) では 80 Hz 変調周波数を採用している。MASTER での 80-Hz ASSR は、FFT を使用したパワースペクトル解析等によって、左右耳それぞれ 4 つの搬送周波数 (Carrier frequency: CF) 500, 1000, 2000, 4000 Hz の反応の閾値を同時に調べる。この 4 つの搬送周波数 (検査周波数) の純音をそれぞれ 70 ~100 Hz の異なる周波数にて振幅変調 (周波数変調も可) した複合 SAM 音 (Sinusoidally Amplitude-Modulated tone) を作成し、左右同時刺激を行う。左右別々の周波数によって振幅変調をかけた合成 SAM 音を聞かせて、蝸牛基板上での各搬送周波数に相当する部位の反応をスペクトル表示して調べる。すなわち得られる反応は脳波上、変調周波数 (Modulation frequency: MF) に一致する周波数成分のパワーが大きくなり、周囲の周波数成分のパワーと F-test で比較して反応の有無を見ることで搬送周波数が聞こえるかどうかを判定する。このようにして、一度に両耳の 4 周波数の聴力閾値が求められる。

40-Hz ASSR の場合

40-Hz ASSR も本来 80-Hz ASSR と同様に蝸牛聴覚フィルタモデルを用いた聴力検査が可能であるが、臨床診断の実用化はされていない。80-Hz ASSR が睡眠中に測定できるのに対して、40-Hz ASSR は覚醒時に現れる反応であり、覚醒時測定は実用的で汎用性が広がるのだが、一方、自発脳波 (spontaneous EEG) の影響が強いため 40-Hz ASSR を用いて聴力検査を行うことは進んでいない。そこで、我々が独自に開発した装置 ([8]) を用

^{*1} 蝸牛は音のパワーと位相の両方を正確に分析して神経信号に変換し、各聴神経線維はある特定範囲の周波数成分だけを伝えるスペクトル分析器であるが、この周波数分析機構を説明するために考えられた仮想の帯域通過フィルタ。即ち、聴神経全体を見ると、それぞれが少しずつ異なった周波数を通させる帯域フィルタの集まりと見なせるが、以下のような特徴を有する。(1) ヒトでは 20 数個の帯域フィルタで構成される。(2) 入力音の周波数や音圧に応じてバンド幅が変化する。(3) 低周波数側と高周波数側で傾斜が異なる (非対称)。(4) 感音難聴者では、帯域幅が広がる。日本聴覚医学会用語集 No.315 より

いて、40-Hz ASSR を測定した．この装置は、MASTER のように複合搬送周波数による測定ではなく、単一搬送周波数 $CF=1000$ Hz に対する脳波を測定する．

計測波形のサンプリング周波数は 1024 Hz、サンプリングポイント 512 点 (500 msec、周波数分解能：2 Hz) を、1 epoch とよぶ．計測脳波は ABR と同様に加算処理が必要であるが、文献 [6] や [4] に基づき (Fig. 2 参照)、epoch 波形を 1 周期 25 ミリ秒 ($= \frac{500}{25 \times 512} = 26$) ポイントシフトして加算する方法を採用した (詳細は [9])．すなわち、計測脳波のサンプリングしたデータ集合

$$D = \{d[t] \mid t = 0, 1, 2, \dots\},$$

に対して、

$$\vec{d}_{m,k} = (d[26(k-1)], d[26(k-1)+1], \dots, d[26(k-1)+511]), \quad m \geq 1, k = 1, \dots, m$$

として、 $M = 10$ epochs の $\vec{d}_{m,k}$ の $k = 1, \dots, m$ についての平均を取った横ベクトルを、40-Hz ASSR の有無を調べる波形データとする．すなわち、

$$\vec{s}_{M,k} = \frac{1}{M} \sum_{k=3}^{m+2} \vec{d}_{m,k}.$$

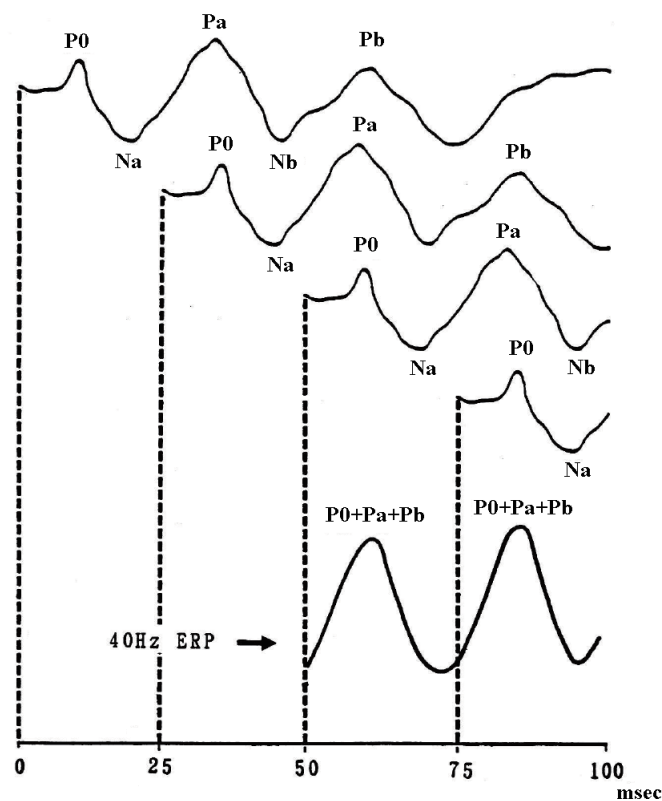


Fig. 2. Relationship of 40-Hz ERP and MLR [4].

3. 1 次元複素連続ウェーブレット解析

本節では, ABR や ASSR 波形に CCWA を適用する. すなわち, ウェーブレット:

$$(3.1) \quad \psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right), \quad a(\neq 0), \quad b \in \mathbb{R}$$

$$(3.2) \quad W_\psi[x(t)] = C(a, b) = \int_{-\infty}^{\infty} x(t) \psi_{a,b}^*(t) dt$$

また, 複素モルレ (complex Morlet) ウェーブレット:

$$(3.3) \quad \psi(t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{t^2}{2\sigma^2}} e^{i\omega_0 t} \quad \text{Fig.3 参照}$$

$$(3.4) \quad W_\psi[x(t)] = C(a, b) = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{2\pi\sigma^2}a} e^{-\frac{(\frac{t-b}{a})^2}{2\sigma^2}} e^{-i\omega_0(\frac{t-b}{a})} dt$$

を ABR および ASSR に適用して, MATLAB 2016a を使って, 観察する. $\sigma = \frac{\sqrt{3}}{2}, \omega_0 = 2\pi(\text{cmorwvf}(\text{Lb}, \text{Ub}, \text{N}, \text{fb}=1.5, \text{fc}=1))$ とした.

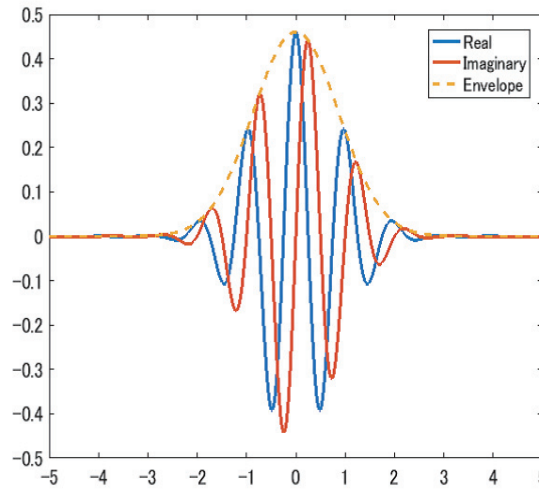


Fig. 3. Complex Morlet wavelet.

3.1 ABR の場合

2000 – average ABR すなわち ABR のサンプリングデータを $x(t)$ として, CCWA を適用した結果例を Fig. 4 に示す. Fig. 4 をみると, ABR は複数の時間-周波数から構成され

ていることがわかる．振幅の大きい波形成分の冗長的な時間-周波数構成を把握することは可能である．

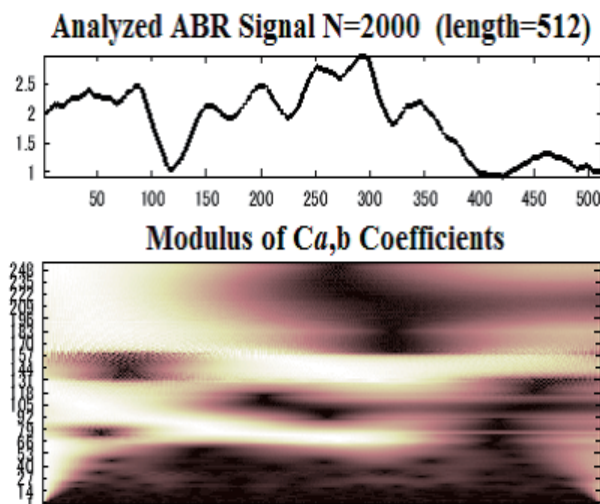


Fig. 4. Complex Morlet wavelet analysis for ABR.

3.2 ASSR の場合

80-Hz ASSR の場合

MASTER により計測された一側性（左耳）難聴症例のデータの波形をサンプリングして CCWA を実施した結果を Fig. 5 に示す．聴力正常の場合は左右耳あわせて 8 つの周波数振幅が観察されるが，この症例では実際，右側の 4 つの周波数のみの反応であった．その結果を Fig. 5 から判読するのは難しい．

40-Hz ASSR の場合

誘発脳波計測後，加算波形 $\vec{s}_{M,k}$ に対して CCWA を実施した．その結果を Fig. 6 に示す．Fig. 6 の左上の波形は同一装置で音刺激を与えずに脳波のみを計測したものである．左側の上から 2, 3, 4 番目の波形はそれぞれ刺激音圧 70dB, 50dB, 30dB nHL の加算波形 $\vec{s}_{M,k}$ である．右側は $\vec{a}_{m,k}$ すなわち，ある epoch 波形である．右最上段は音刺激がない場合，上から 2, 3, 4 番目は刺激音圧 70dB, 50dB, 30dB nHL の epoch 波形である．この図からは次のことが得られた：

1. 音刺激音圧あり加算波形 $\vec{s}_{M,k}$ では，40-Hz ASSR の顕著な特徴が得られた．
2. 音刺激音圧あり加算波形の振幅は，音圧の強度に影響されず，一定であった．
3. 音刺激音圧なし加算波形でも，40 Hz 付近の反応を多少観察できる（ガンマ波?）．
4. 音刺激音圧なし epoch 波形では，40 Hz 付近の反応は観察できない．
5. 音刺激音圧あり epoch 波形の振幅は，音圧の強度の影響をあまり受けていない．

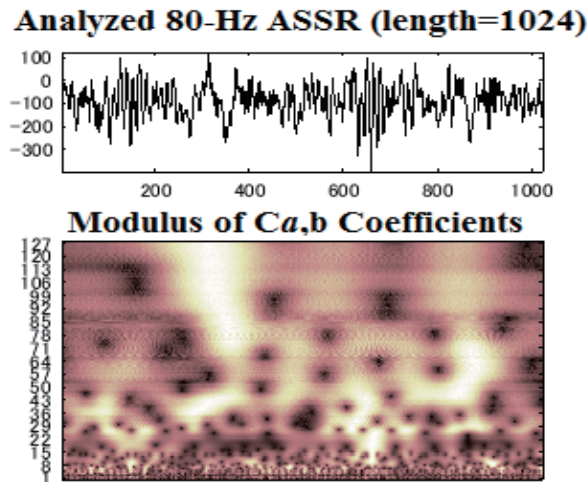


Fig. 5. Complex Morlet wavelet analysis for 80-Hz ASSR (MASTER data).

4. 1次元離散定常ウェーブレット解析

SWT のアルゴリズムは次のとおりである：

1. 長さ N の信号 s が与えられ，SWT の第 1 ステップは，近似係数 (approximation coefficients) cA_1 ，詳細係数 (detail coefficients) cD_1 の 2 つの係数集合からはじまる．
2. これらのベクトルは，approximation に対するローパスフィルタ Lo.D，detail に対するハイパスフィルタ Hi.D について s のコンボリューションによって得られる．
3. ただし， cA_1 ， cD_1 の長さはどちらも N である．
4. 次のステップは，近似係数 cA_1 から，前のステップで用いた s を cA_1 に置き換えたフィルタを用いてアップサンプリングして，新たなフィルタを得ることで上記同様に 2 つのパートを作成する．
5. SWT によって， cA_2 ， cD_2 が作成される．
6. この処理を繰り返す (Fig. 7 参照)．

SWT は，特殊な直交ウェーブレットあるいは特殊な直交ウェーブレット分解フィルタを用いてマルチレベル 1 次元定常ウェーブレット分解をおこなう．波形再構成には ISWT (Inverse discrete stationary wavelet transform) を用いる．ISWT は，特殊な直交ウェーブレットあるいは特殊な直交ウェーブレット再構成フィルタ (Lo.R and Hi.R) を用いてマルチレベル 1 次元定常ウェーブレット再構成をおこなう．ウェーブレット関数は再構成可能な双直交ウェーブレットである Bi-orthogonal ウェーブレット (Bior. 5.5) を用いた (Fig. 8)．

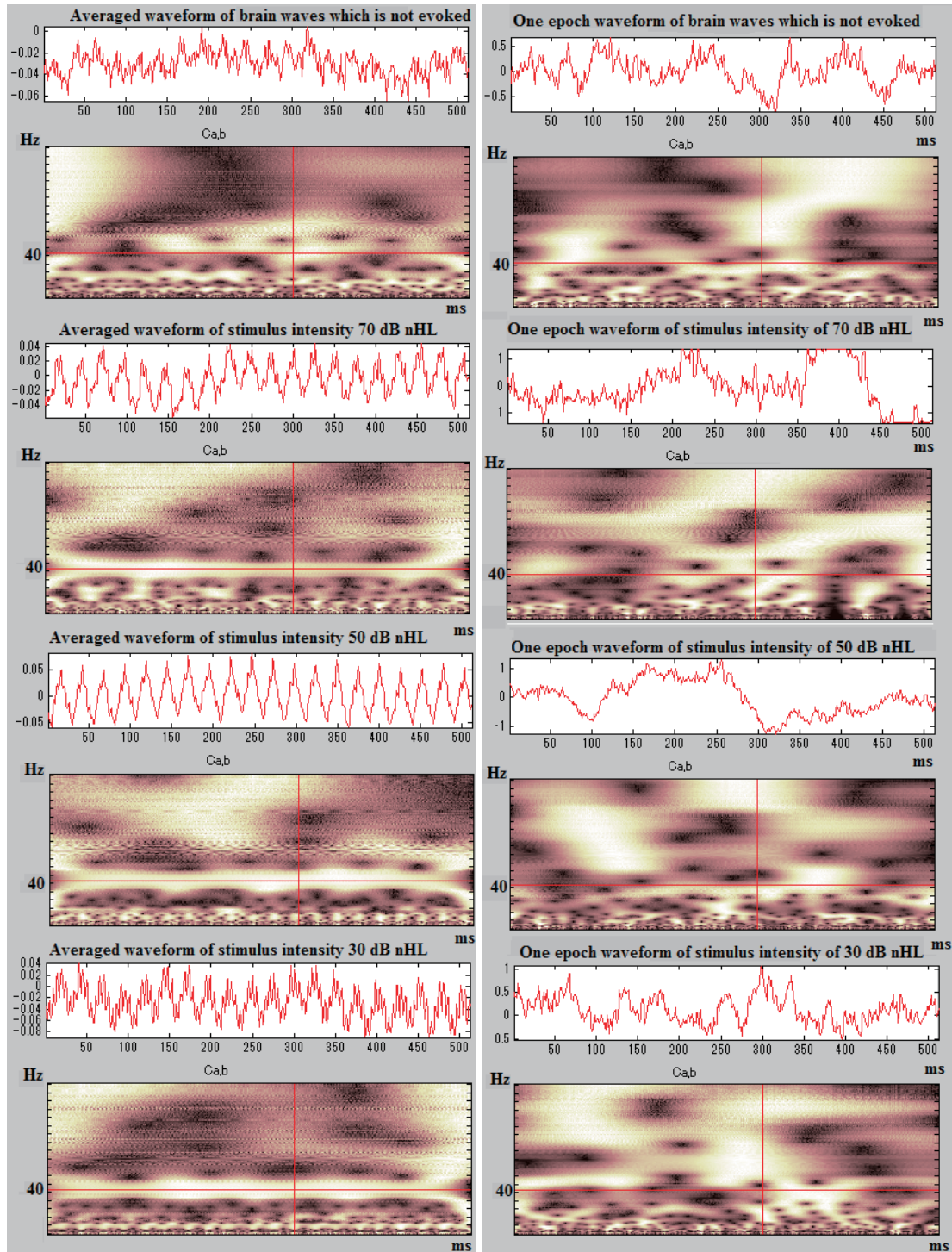


Fig. 6. CCWA to brain waveforms.

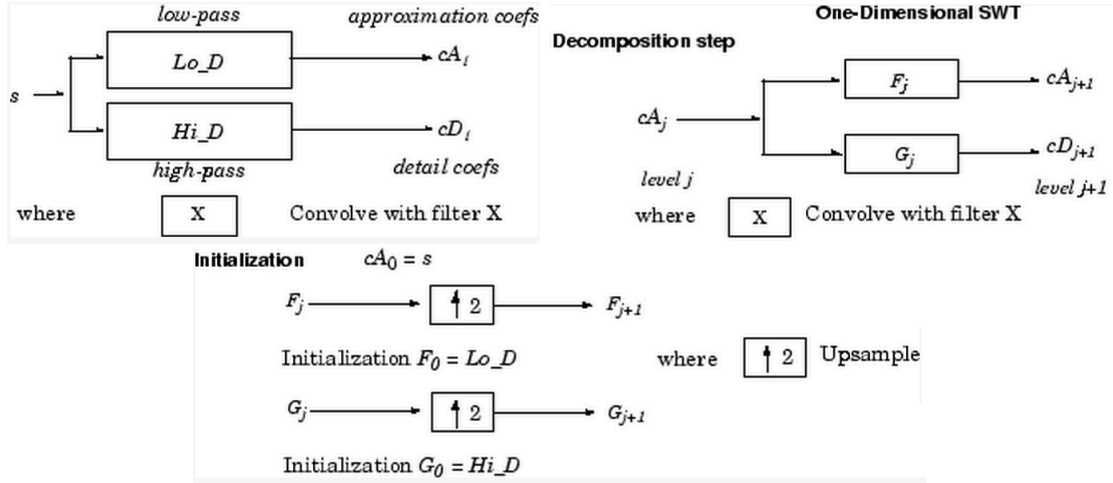


Fig. 7. SWT procedures.

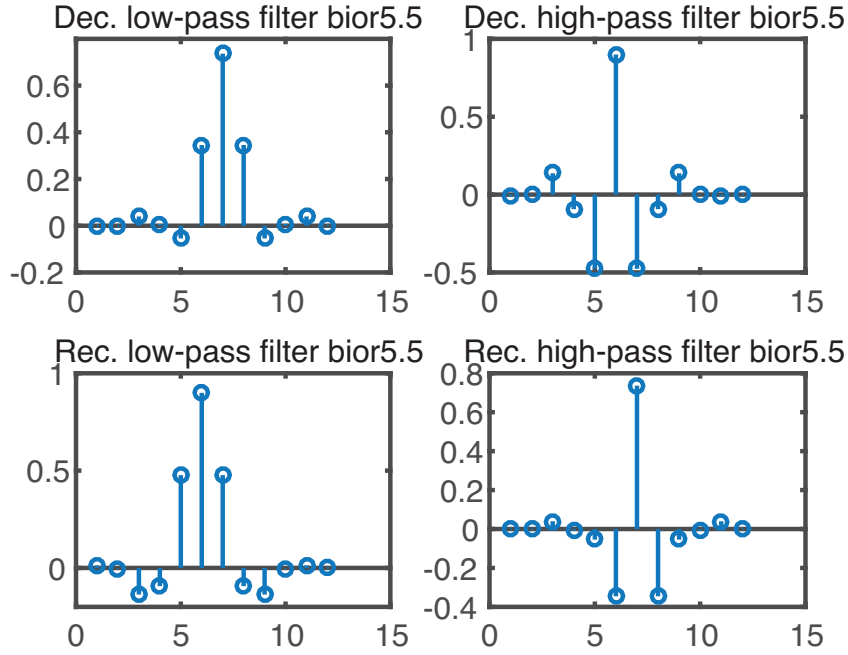


Fig. 8. Bior 5.5 filters.

4.1 ABR の場合

40-Hz ASSR について例えば文献 [10] および [11] で述べている．本稿では ABR に SWT を実施した場合について取り上げる（[12]–[16]）．我々の実験による計測波形のサンプリング周波数から得られる分解レベルの構成周波数を Table 1 に示す．

定義 4.1 低周波数領域 (A8+D8 の構成周波数) を緩徐波成分 (slow ABR)，高周波数領域

Table 1. Decomposition levels of SWT and each frequency band.

Decomposition level	Frequency band
D1	12500 - 25000 Hz
D2	6250 - 12500 Hz
D3	3125 - 6250 Hz
D4	1562 - 3125 Hz
D5	781 - 1562 Hz
D6	390 - 781 Hz
D7	195 - 390 Hz
D8	97 - 195 Hz
A8	0 - 97 Hz

Result 1 : Decomposition levels of Overlapped averaging waveforms

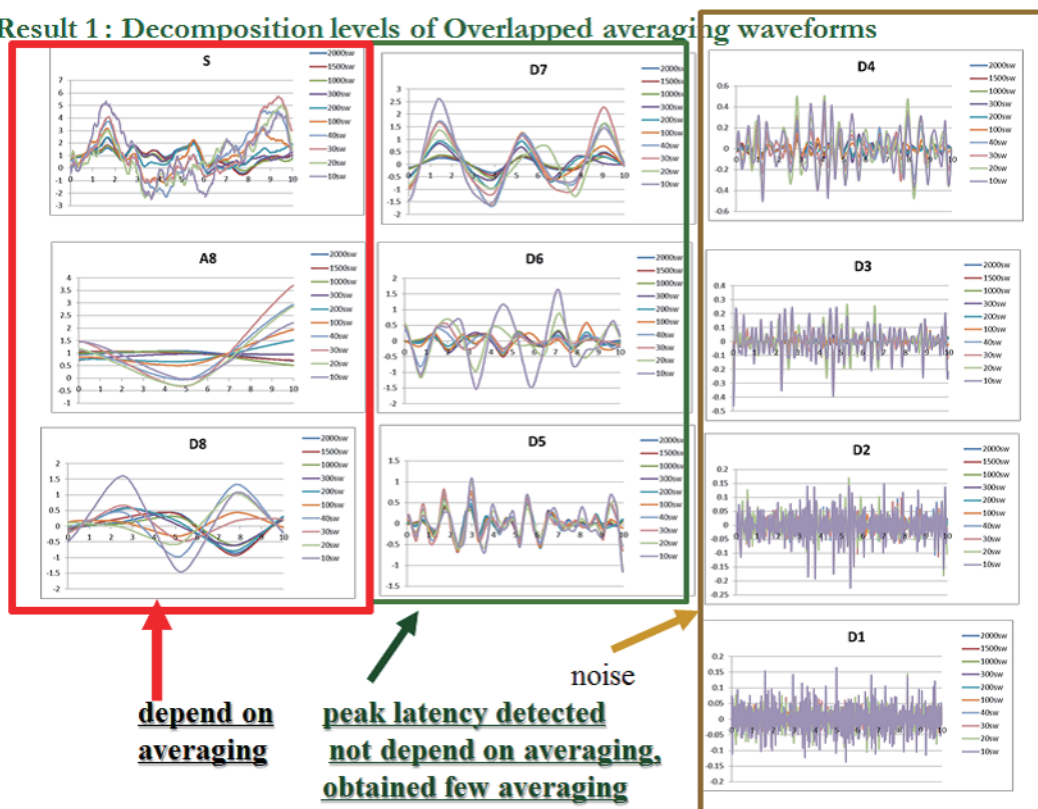


Fig. 9. SWT for ABR.

($D1+D2+D3+D4$ の構成周波数) をノイズ成分, 両者以外の周波数領域 ($D5+D6+D7$ の構成周波数) を速波成分 (fast ABR) とよぶ.

Fig. 9 は, $ABR_N, N = 10, 20, 30, 40, 100, 200, 300, 1000, 1500, 2000$ の $Di_N, i = 1 \text{ to } 7$ と $Ai_N, i = 8$ 波形の加算ごとの重ね書きを示す. 加算によって振幅の多少の減少があるが, 一方, $Di_N, i = 5 \text{ to } 7$ 各波のピーク潜時はきわめて少ない加算回数から観察される. 特に $D5_N$ が顕著で $N = 10$ 以降, 潜時 (位相) がほとんど変わらない. 波形相関係数も 0.9 であった. これを **locked phase** ということにする.

一方, slow ABR すなわち, $A8_N, D8_N$ では, 加算によって波形の **synchronized phase** を観察した. **Sine Wave Fit アルゴリズム**を用いて波形を近似することや, 蔵本モデル (例えば文献 [17] 参照) の応用を試みている.

5. まとめと考察

ABR のピーク潜時の特徴は比較的わかっていることが多いが, 一方, 構成周波数については, 自明な単一周波数で構成されていないことがわかっている. ABR の反応の有無の判定ではピーク潜時と周波数情報の両方を同時に必要とする. 従って, ABR 波形の抽出および同定において, SWT による時間-周波数解析は非常に効果的であることがわかった.

特に, 加算過程の波形 ABR_N に SWT による時間-周波数解析を適用すると加算の早い段階で反応が得られる fast ABR と位相同期がカギとなる slow ABR の振る舞いを詳しく調べ, 自動判定や数式によるモデル化によってその機能が明快になりつつある. このモデル化の生理学・解剖学的知見との照合も重要であるが, 従来法よりも高速かつ詳細な聴性誘発脳波解析が可能となった. 一方, slow ABR, $A8$ は, 振幅において大きな影響があるレベルだが, 加算 300 回程度すなわち, 3 秒程度に短縮が可能である. なによりもモデル波形を参照することで, 反応が得られない場合に必要以上に加算しなくても, 反応なしの判定の確度が増すと考える.

これら SWT による ABR の詳細分析を受けて, ASSR について考察すると, ASSR は ABR の分解レベル $A8$ の影響を強く受ける反応であるので, 一定の加算同期が必要であること, また, 波形振幅は加算によってあまり影響がでないことが判明した. 単一搬送周波数刺激を用いる場合は, CCWA によって時間-周波数に関する顕著な特徴が得られた.

ウェーブレット解析について, 総当たり法を用いて, 要求の結果が得られるものを応用したが, 理論的な妥当性や根拠について, より数学的な立証について議論することあるいはより良い方法の提案をしていただくことを望む. その結果, ウェーブレット解析の聴性誘発脳波への応用の精度向上をめざしたい.

謝辞 実験の際, 千葉大学 CFME の支援を受けた. また, 科研費 (C)26400199 の支援を受けたことに感謝する.

参考文献

- [1] 山田道夫, 萬代武史, 芦野隆一, 応用のためのウェーブレット, シリーズ応用数理第 5 巻, 共立出版, 2016.
- [2] 入野俊夫, はじめての聴覚フィルタ, ASJ tutorial, 日本音響学会誌, 66 巻 10 号, 506–512, 2010.
- [3] 日本聴覚医学会編, 聴覚検査の実際 改訂 2 版, 南山堂, 2004.
- [4] 市川銀一郎編, 初心者のための聴性誘発反応アトラス, 広川書店, 1989.
- [5] 青柳優, 聴性定常反応 その解析法・臨床応用と起源, リオン株式会社, 2005.
- [6] R. Galambos et al, A 40-Hz auditory potential recorded from the human scalp, Proc. Nati. Acad. Sci. USA, **78**(4), 2643–2647, 1981.
- [7] M. S. John, et al., MASTER: a Window program for recording multiple auditory steady-state responses, Comput. Methods Programs Biomed., **61**, 125–150, 1998.
- [8] 井川他, 聴性定常反応による聴力検査装置の試作 —PXI-4461 による計測精度の向上—, 千葉大学 CFME, 2009.
- [9] 井川, 守本, 芦野, 40Hz 聴性定常反応の加算法と離散定常ウェーブレット解析, JSIAM 年会, 47–48, 2011.
- [10] N. Ikawa, A. Morimoto and R. Ashino, Waveform analysis of 40-Hz auditory steady-state response using wavelet analysis, IEEE conference of Wavelet Analysis and Pattern Recognition (ICWAPR), 397–402, 2012.
- [11] N. Ikawa, A. Morimoto and R. Ashino, An application of wavelet analysis to procedure of averaging waveform of 40-Hz auditory steady-state response, IEEE conference of Wavelet Analysis and Pattern Recognition (ICWAPR), 79–84, 2013.
- [12] 井川, 聴性脳幹反応加算時間経過波形のウェーブレット変換による再構成波形の特徴とモデル化について, 数理解析研究所講究録 **1972**(2015), 23–41.
- [13] N. Ikawa, A. Morimoto, and R. Ashino, The detection of the relation of the stimulus intensity-latency of auditory brainstem response using optimal wavelet analysis, in the proceedings of ICWAPR2014(2014), 127–133.
- [14] N. Ikawa, A. Morimoto, and R. Ashino, A phase synchronization model between auditory brainstem response and electroencephalogram using the reconstructed waveform of multi-resolution discrete stationary wavelet analysis, in the proceedings of ICWAPR2015(2015), 111–116.
- [15] 井川信子, 守本晃, 芦野隆一, 加算波形のウェーブレット解析による聴性脳幹反応のモ

デル化, 応用数理学会 2015 年度年会予稿集, 2015.9.

- [16] N. Ikawa, A. Morimoto, and R. Ashino, Optimum wavelet filter estimating peak latencies of Auditory Brainstem Response waveform, in the proceedings of ICWAPR2016(2016), 189–194.
- [17] 蔵本由紀, 河村洋史, 同期現象の数理: 位相記述によるアプローチ, 培風館, 東京, 2010.

井川 信子 (流通経済大学法学部)

〒301-8555 茨城県龍ヶ崎市 120

E-mail: ikawa@rku.ac.jp

守本 晃 (大阪教育大学教育学部)

〒582-8582 大阪府柏原市旭ヶ丘 4-698-1

E-mail: morimoto@cc.osaka-kyoiku.ac.jp

芦野 隆一 (大阪教育大学教育学部)

〒582-8582 大阪府柏原市旭ヶ丘 4-698-1

E-mail: ashino@cc.osaka-kyoiku.ac.jp

Multi-Dimensional Discrete Wavelet Transformation and Its Application to Polarimetric SAR Classification

Kohei Arai

Graduate School of Science and Engineering, Saga University, 1 Honjo,

Saga 840-8502 Japan

arai@is.saga-u.ac.jp

Abstract. Discrete Wavelet Transformation: DWT is explained in a comprehensive manner starting from one dimensional to multi-dimensional DWT. One of the applications of the multi-dimensional DWT is introduced for polarimetric Synthetic Aperture Radar: SAR classification of sea ice type discriminations. Polarimetric SAR onboard remote sensing satellite allows sea ice monitoring as well as land use classifications. One of the specific features of polarimetric SAR is to clarify scattering properties of the ground cover targets. By using the differences of the scattering properties, land use classifications as well as sea ice discriminations can be done. In order to extract scattering properties, scattering matrix is derived from the polarimetric SAR data. From the scattering matrix, trajectory in a phase space is calculated. On the other hand, time and frequency analysis is available by the multi-dimensional DWT. In order to use the specific feature of the trajectory in the phase space, behavior of the scattering properties, multi-dimensional DWT is applied to the trajectory. Thus the specific polarimetric features of the designated ground cover targets are extracted and classified using supervised classifications.

Keywords: Wavelet, Multi-Resolution Analysis, Polarimetric SAR, Classification

1 Introduction

Radar polarimetry allows measurement the physical characteristics such as di-electric constant, slope of the ground cover targets as well as directionality of artificial objects by using scattering mechanism between electromagnetic (EM) wave and the targets [1],[2]. Polarimetric SAR image classification with the following three components of the polrimetric SAR data, (1) transmit Electro-magnetic wave with Horizontal Polarization(H-Pol) and receive the echo from the ground with H-Pol(HH), (2) transmit Electro-magnetic wave with H-Pol and receive the echo with Vertical Polarization(V-Pol)(HV) and (3) transmit Electro-magnetic wave with V-Pol and receive the echo with V-pol(VV) is widely available [3],[4]. On the other hand, the extraction of the scattering characteristics of the targets of interest by applying eigen value decomposition to the covariance matrix derived from the scattering matrix which is calculated from the three components are proposed [5]. Furthermore, the classification methods with the single / double /multiple, odd

/ even / diffuse, and odd / even / Bragg / multiple scattering components derived from the eigen value decomposition were proposed [6] while the classification methods with the sphere / deplane / helix, and sphere / Bragg / double of scattering components which are based on the spherical polarization which are derived from the scattering matrix were also proposed [7],[8]. Aforementioned proposed methods were reviewed [9]. Moreover, the classification method with the entropy (H) which is defined with the sum of the first to third eigen values and the ratio of each eigen values, the anisotropy (A) which is defined as the ratio of sum and subtraction of the second and the third eigen values and cosine α ($\cos(\alpha)$) which is defined with the elements of the eigen vector corresponding to the first eigen value which is called coherency matrix (3 by 3) was proposed by E.Pottier [10]. The application of these methods to sea ice discrimination (such as thin ice (TI), smooth first year ice (SF), rough first year ice (RF) and open water (OW)) with the polarimetric SAR were attempted by using H, A, and $\cos(\alpha)$ [11]. Classification performance, however, were not satisfactory (20-40% of classification errors were occurred for the classification of sea ice into four classes, ridged, compressed, new forming and smooth surface due to the fact that scattering mechanism based features were not used effectively. Meanwhile polarimetric SAR image classification with polarization signature which are derived from Stokes or Muller or scattering matrix is widely available [12]. Polarization signature represents the scattering mechanism, in particular, surface roughness of the targets in concern. One of the problems on the classification with polarization signature is classification performance. The method for effective utilization of polarization signature is still unclear to improve classification performance. The method proposed here is for extraction of effective information from the polarization signature by transforming the polarization signature onto an eigen space (eigen value decomposition). As the results from the eigen value decomposition which corresponds to the largest eigen value, a trajectory can be drawn. The trajectory represents the scattering mechanism in concern so that the largest curvature of the trajectory represents the most effective representatives of the scattering mechanism of the target of interest [13]. This is the theoretical background to propose the utilization of maximum curvature of the trajectory in an eigen space which is derived from the polarization signature to the sea ice classification [14], [15].

Polarization signature is the back scattering cross section in the space of the ellipticity angle and the orientation of the electric field vector. It represents the scattering mechanism, in particular, surface roughness of the targets. Polarization signature is useful for visual perception of the scattering characteristics of the targets. It, however, is not easy to utilize for classifications. One of the problems of the classification with polarization signature is poor classification performance. A method for utilization of the polarization signature is proposed here. Also, wavelet MRA: Multi Resolution Analysis [16] – [18] is applied to the polarimetric SAR imagery data for enhancing high frequency components of the data for improving classification performance.

Firstly, the proposed method is introduced with a theoretical background followed by experimental data and the results from the experimental are described together with the results from a comparative study between

the proposed method and the existing conventional methods. Finally, concluding and remarks are described with some discussions.

2 Proposed Method

2.1 Full Polarimetric SAR

Received echo signal $(E_h, E_v)^T_{\text{rec}}$ is expressed as follows,

$$\begin{pmatrix} E_h \\ E_v \end{pmatrix}_{\text{rec}} = \frac{e^{-jkR}}{kR} \begin{pmatrix} S_{hh} & S_{hv} \\ S_{vh} & S_{vv} \end{pmatrix} \begin{pmatrix} E_h \\ E_v \end{pmatrix}_{\text{ill}} \quad (1)$$

where S is a scattering matrix composed with HH, HV, VH and VV components of S_{hh} , S_{hv} , S_{vh} and S_{vv} , $(E_h, E_v)^T_{\text{ill}}$ denotes incident Electric Magnetic: EM wave, k is wave number of the incident EM wave, R denotes range, respectively. From these elements of S , Muller matrix is calculated while Stokes vector at the receiver J_r is calculated with $(E_h, E_v)^T_{\text{rec}}$ and also Stokes vector at the transmitter J_t is calculated with $(E_h, E_v)^T_{\text{ill}}$ thus the polarization signature σ^0 is represented by the following equation,

$$\sigma^0 = C J_r^T \langle M_s \rangle J_t = \sigma^0(\chi_r, \phi_r, \chi_t, \phi_t) \quad (2)$$

$C : \text{const.}$

where the χ and ϕ denote the ellipticity angle and the orientation angle of the orientation of the electric field vector, respectively of which the polarization is described with the two parameters. The polarization signature describes the scattering coefficient as a function of any assumed transmit and receive antenna polarization and allows measure the variation of the scattering coefficient with polarization so that the different targets show the different polarization signature. The method for effective utilization of polarization signature is still unclear to improve classification performance.

Let X be matrix with three scattering components as follows,

$$X = [S_{hh} S_{hv} S_{vv}]^T, C = \langle X, X^{*T} \rangle \quad (3)$$

$$C = \lambda_1 K_1 (K_1^*)^T + \lambda_2 K_2 (K_2^*)^T + \lambda_3 K_3 (K_3^*)^T$$

where C denotes covariance matrix of X and $\lambda_1, \lambda_2, \lambda_3$ are eigen values of C . These are corresponding to odd times scattering, even times scattering and diffuse scattering, respectively. Also $*$ and T denotes complex conjugate and transpose, respectively. Thus,

$$\frac{\lambda_i}{\lambda_1 + \lambda_2 + \lambda_3}, (i = 1, 2, 3)$$

is called contribution factor. If the equation (4) is formulated,

$$\begin{aligned}
S_{RR} &= \frac{1}{2}(S_{hh} - S_{vv} + j2S_{hv}) \\
S_{LL} &= \frac{1}{2}(S_{hh} - S_{vv} - j2S_{hv}) \\
S_{LR} &= \frac{1}{2}(S_{hh} + S_{vv})
\end{aligned} \tag{4}$$

then spherical, di-plane and helix components are defined as follows,

$$K_s = |S_{LR}|, K_d = |S_{LL}|, K_h = |S_{RR}| - |S_{LL}|, \text{ for } |S_{RR}| > |S_{LL}| \tag{5}$$

Thus contribution factor for each is defined as follows,

$$\frac{K_i}{K_s + K_d + K_h}, (i = s, d, h) \tag{6}$$

Polarization signature represents the polarization feature of the ground cover target, obviously.

The polarization signature describes the scattering coefficient as a function of any assumed transmit and receive antenna polarization and allows measure the variation of the scattering coefficient with polarization so that the different targets show the different polarization signature as shown in Figure.1.

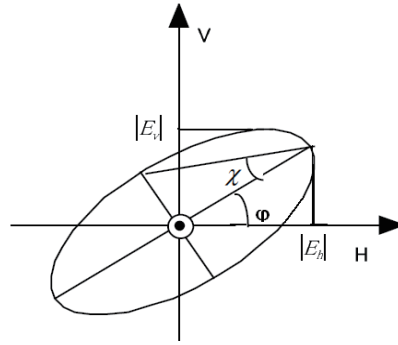


Figure 1. Definition of polarization signature

The method for effective utilization of polarization signature is still unclear to improve classification performance. Once the scattering matrix which is composed with HH, HV, VH and VV components of S_{hh} , S_{hv} , S_{vh} and S_{vv} , are calculated from fully polarimetric SAR data, then the back scattering cross section σ^0 at the arbitrary polarization angle which ranges from HH (zero polarization angle) to VV (90 degrees of polarization angle) for cross polarized data as well as co-polarized data is calculated (the distance among the classes for cross polarization is greater than that of co-polarization, in general). It is called polarization signature. If some quantization steps, q ($q=5$ degree in this case) is set to the polarization angle, then $90/q$ dimensional data is reduced. After that such a multi-dimensional data is projected in eigen space through principal component analysis. Thus the polarization signature becomes a trajectory in the eigen space. The trajectory represents scattering mechanism of the target. One of the features of the trajectory is maximum curvature. An osculating circle is defined at each data point (every 5 degree of the polarization angle, in this case) of the trajectory. Then the curvature K at the data point is calculated in accordance with the following definition.

$$K = 1/r \quad (7)$$

where r is radius of the osculating circle. The maximum curvature is defined as the maximum K . The proposed classification method is based on the maximum likelihood classification with the received signals of three different polarizations, (HH and VV of co-polarization) and HV of cross-polarization together with the maximum curvature.

2.2 Classification Method

Classification method proposed here is based on Maximum Likelihood Classification using not only polarimetric SAR radar echo signals, but also wavelet frequency components as features for classification. In particular, high frequency components derived from wavelet Multi Resolution Analysis: MRA is used as spatial features of the target class in concern.

2.3 Wavelet Multi-Resolution Analysis

One dimensional wavelet transformation is expressed with the following equation.

$$F = C_n f \quad (8)$$

where F, f denotes wavelet frequency component and radar echo signal as a function of time. C_n denotes wavelet transformation matrix which is expressed as a bi-orthogonal function based on base functions. C_n can be determined with a reference to the appendix. Therefore,

$$C_n C_n^T = I \quad (9)$$

Then, f is converted to

$$F_1 = (L_1, H_1), F_2 = C_n L_1 = (L_2, H_2), F_3 = C_n L_2 = (L_3, H_3), F_m = C_n L_{m-1} = (L_m, H_m) \quad (10)$$

Also f is reconstructed as

$$C_n^{-1} F_m = C_n^{-1} (L_m, H_m) = L_m^{-1}, \dots, C_n^{-1} F_2 = L_1, C_n^{-1} F_1 = f \quad (11)$$

The suffix of 1 to m is called “level”. Level m implies that wavelet transformation is applied m times. MRA ensure that the original signal can be reconstructed with the wavelet coefficients or frequency components of the level m . The frequency components derived from MRA are corresponding to the level m . Therefore, MRA does work as a filter bank.

There are some base functions such as Haar, Daubechies¹, etc. Through the experiment with preliminary simulation of radar echo data, Daubechies base function is selected. Daubechies base function is one of bi-orthogonal functions. MRA is applied to the polarization signature then the extracted high frequency component is added to the typical features of received polarization power signals, HH, HV, and VV components where HH denotes the horizontal polarization of electro-magnetic wave is transmitted then the returned echo signal is received in horizontal polarization. Thus, spatial feature of the polarization of feature of the ground cover target is taken into account in the classifications. Typically, three polarization power signals are acquired with polarimetric SAR system.

2.4 Methods for Comparison

The classification performance of the proposed method is compared to the Maximum Likelihood classification with only three components of the polarimetric SAR data of HH, HV, VV, with the three components + H + A + cos(α), with the three components + Cl, and with the three components + Odd + Even + Diffuse for discrimination among the following three classes, Urban, Vegetation, and Paddy field. Also, classification performance is compared with the previously proposed classification method with information from the polarization signature by using eigen value decomposition of the polarization signature. As the results from the eigen value decomposition which corresponds to the largest eigen value, a trajectory can be drawn. The trajectory represents the scattering mechanism in concern so that the largest curvature (Cl) of the trajectory represents the most effective representatives of the scattering mechanism of the target of interest. If the largest curvature is large, then the polarization signature is steeply while the largest curvature is small, then the polarization signature is calm. The proposed classification method is based on the well known maximum likelihood classification with the received signal of the three different polarizations, co-polarization (HH and VV) and cross-polarization (HV) as well as the maximum curvature of the trajectory in the eigen space through the eigen value decomposition from the polarization signature.

3 Experiments

3.1 Data Used

The PI-SAR (Polarimetric and Interferometric SAR) data of Tsukuba in Japan which was acquired by CRL (Communication Research Laboratory, current NICT: National Institute of Communication Technology) and NASDA (National Space Development Agency of Japan, current JAXA: Japan Aeronautics Exploration

¹ Daubechies base function is defined as $\{\alpha\}$ satisfying the following conditions,

$$\begin{aligned}\phi(x) &= \sum_k \alpha_k \sqrt{2} \phi(2x - k) \\ \beta_k &= (-1)^k \alpha_{1-k} \\ \varphi(x) &= \sum_k \beta_k \sqrt{2} \phi(2x - k)\end{aligned}$$

Agency) was used. Tsukuba, Ibaraki prefecture in Japan which consists of urban, forest and paddy fields of September 30 1997 were used. The major characteristic of the PI-SAR is shown in Table 1.

TABLE 1. MAJOR CHARACTERISTICS OF PI-SAR

Instrument	NASDA/L-band SAR
Center Frequency	1.27GHz
Peak power	3.5KW
Bandwidth	50MHz
Antenna size	1.6mx0.7m
Polarization	HH/HV/VH/VV (Full polarization)
Incidence angle	20-60 degrees(Fixed)
Swath width	42.5km
Spatial resolution	3m
Quantization bit	8bits(I and Q signals)

3.2 Polarization Signature

From the data of the SSC: Single-look Slant-range Complex, 5km by 5km of the area was extracted as is shown in Figure.2. The image is 8 look processed HH, HV and VV polarization of data which consists of 1000 by 987 pixels. One pixel is corresponding to 3 by 3 meters. Square boxes of blue, yellow, and green show the training area of Paddy fields (After harvest), Urban areas, and Vegetated areas, respectively.

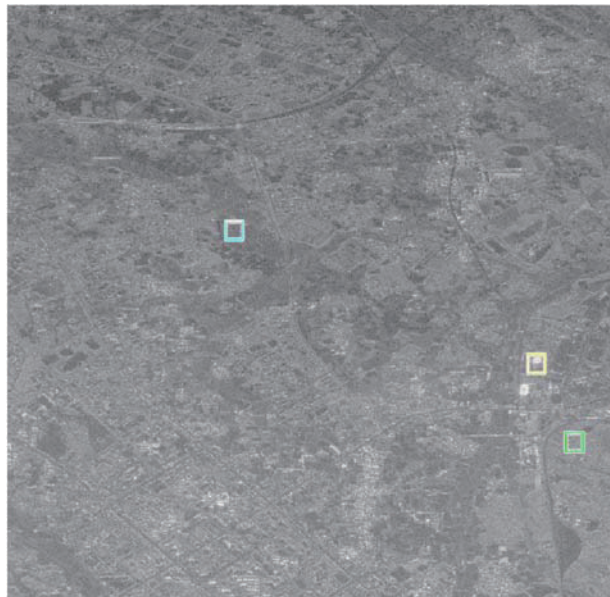


Figure 2 Polarimetric SAR image (SSC) of Tsukuba Acquired on September 30 1997

Using equation (1), polarization signature can be calculated by pixel by pixel because the parallel polarization, Co-pol. (HH, VV) and cross polarization, Cross-pol. (HV, VH) are acquired by pixel by pixel. Figure.3 shows the calculated polarization signature for each class. The calculated polarization signatures are different each other. Therefore, it can be classified. The problem of the conventional classification method which uses the polarization signature and the returned echo signal powers, HH, HV, and VV is poor classification performance. The proposed method utilizes high wavelet frequency components other than these polarization signature and returned echo signal powers.

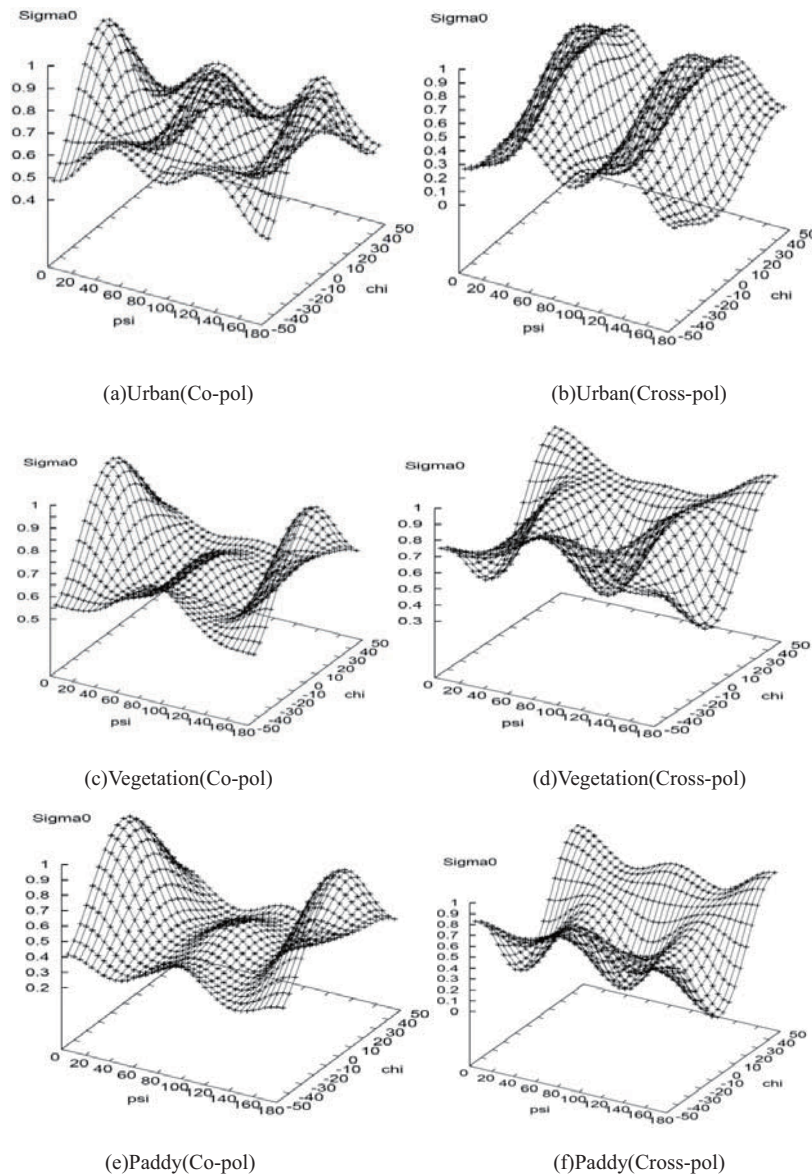
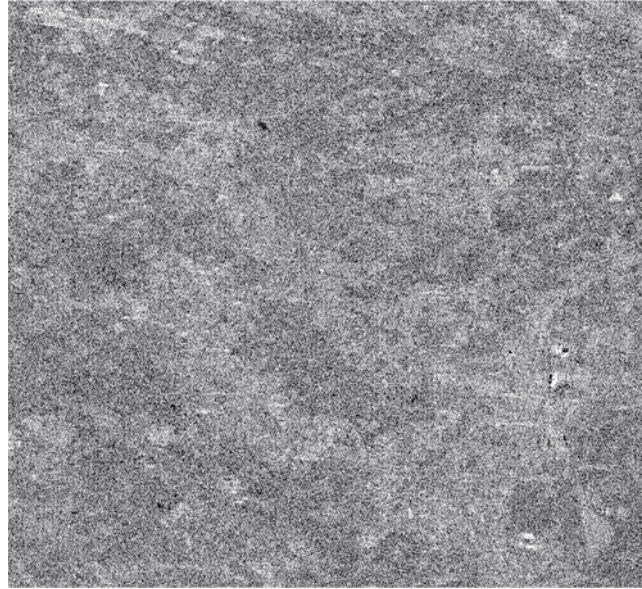


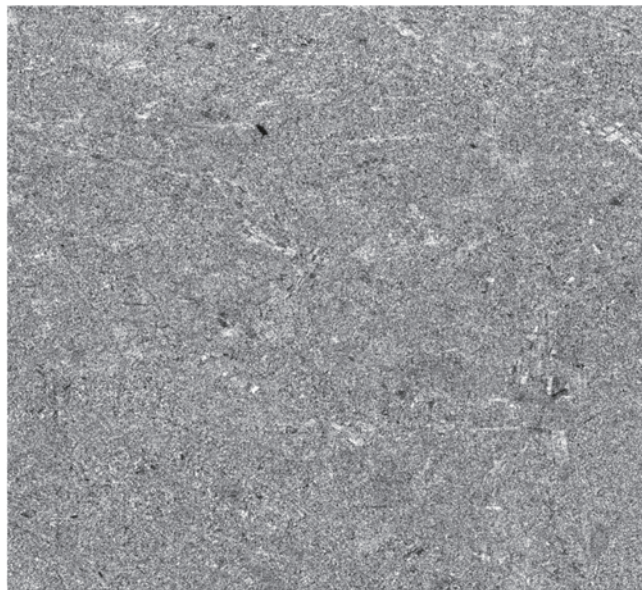
Figure 3. Examples of the calculated polarization signatures of the classes, Urban, Vegetation, and Paddy extracted from the 16 by 16 pixels of square areas which are corresponding to the designated classes in concern

Figure.4 (a) and (b) shows examples of the high frequency component (HH₄: Level 4 of DWT with Daubechies base function) of the polarization signature of image which is shown in Figure.3, and the

maximum curvature of the trajectory in eigen vector space which is converted from the polarization signature image, respectively.



(a) Example of the high frequency component of HH_4 of the polarization signature of image which is shown in Figure.3



(b) Maximum curvature image of the trajectory in eigen vector space which is converted from the polarization signature image

Figure.4. Images of the features of the previously proposed maximum curvature of the trajectory in eigen space which is converted from the polarization signature and the high wavelet frequency component image of the polarization signature.

3.3 Class Features

Because the PI-SAR frequency is 1.27 GHz, the contribution of leaves of vegetation is very weak (penetrated) while the contribution from the ground surface and the relatively large trees is major. Therefore, polarization signature of vegetated areas is similar to the paddy fields after harvest except surface flatness. On the other hand, HH return echo signal power and polarization signature for both urban areas and vegetated areas are similar. As the results from the aforementioned reasons, all these three classes are difficult to classify. It is confirmed that three components of return echo signal power, HH, HV, and VV utilized classification makes about 80 % of classification performance. In order to improve classification performance, the following feature is proposed by the author. That is maximum curvature of the trajectory in the eigen vector space which is converted from the polarization signature with arbitrary polarization angle response with 5 degree intervals. The proposed feature does work for classification for the classes with different surface roughness, in particular. Although it is effective to improve classification performance, it is complicated and requires computation resources. The proposed classification method is based on wavelet frequency components of the calculated polarization signature. Therefore not significant computation resources are not required and it is quite simple.

It is difficult to discriminate between urban and vegetation classes because the previously proposed maximum curvature of urban and vegetation is similar as shown in Figure.4 (b). On the other hand, the proposed feature of high frequency component HH₄ of image shows clear difference between urban and vegetation classes. Therefore, HH₄ frequency component which is derived from the level 4 of MRA output, HH₄, HL₄, LH₄, and LL₄ of the image which consists of 32 by 32 pixels with 5.6 degree step of the polarization angle of the each dimension of polarization signature is used for classification.

3.4 Classification Performance

Using training samples extracted from the training areas which are indicated in Figure.2, training performance is evaluated. A comparison is made for the following three Maximum Likelihood based classification method with (1) received three echo signal power, HH, HV, VV, with (2) three power and the previously proposed maximum curvature, and with (3) three power and the HH₄ of each dimension of polarization signature. Classification performance is evaluated with confusion matrix and the k statistics which is shown in the following equation (12),

$$k = \frac{P(A) - P(E)}{1 - P(E)} \quad (12)$$

where $P(A)$, $P(E)$ denotes correct classification probability and classification error probability. Table 2, 3, 4 shows the results of classification performance.

TABLE 2. THREE RECEIVED ECHO SIGNAL POWER ONLY (K=89.05)

	Urban	Vegetation	Paddy
Urban	75.2	24.8	0.0
Vegetation	2.4	97.6	0.0
Paddy	0.8	1.6	97.6

TABLE 3. THREE POWER AND THE MAXIMUM CURVATURE (K=95.62)

	Urban	Vegetation	Paddy
Urban	89.8	10.2	0.0
Vegetation	1.2	98.8	0.0
Paddy	0.4	0.8	98.8

 TABLE 4. THREE POWER AND HH₄ (K=98.0)

	Urban	Vegetation	Paddy
Urban	94.9	5.1	0.0
Vegetation	0.0	100.0	0.0
Paddy	0.0	0.8	99.2

As is shown in the previous section, it is difficult to discriminate between urban and vegetation classes if only three power of features are used. It is improved by adding the features of maximum curvature and HH₄, remarkably. On the other hand, k -statistics shows that 7.38% of improvement is achieved by adding the maximum curvature and 10.05% improvement is achieved by adding HH₄.

Conclusion

It is found that the proposed features of three received echo signal power, HH, HV, and VV as well as HH₄ of high frequency component derived from MRA analysis with polarization signature are effective to improve classification performance for PI-SAR types of polarimetric SAR images. More than 10% of improvement of the classification performance is confirmed by adding HH₄. The improvement depends on the level of MRA, because there is the most appropriate level depending on the frequency component of the polarization signature of the ground cover target. The computational resource requirement for the previously proposed maximum curvature of the trajectory in eigen vector space which is converted from the polarization signature is twice much larger than that of the proposed MRA based feature utilized classification. Thus it may said that the proposed high frequency component of polarization signature based classification method is superior to the conventional classification method with three echo signal powers, and to the previously proposed maximum curvature based classification method in terms of classification performance and the required computation resources.

Appendix

(1) Another polarization signatures of asphalt and water body

Other than vegetation, urban, and rice paddy field, other polarization signatures are shown in Figure A1. Sigma note, σ^0 of asphalt is relatively high while that of water body is quite low.

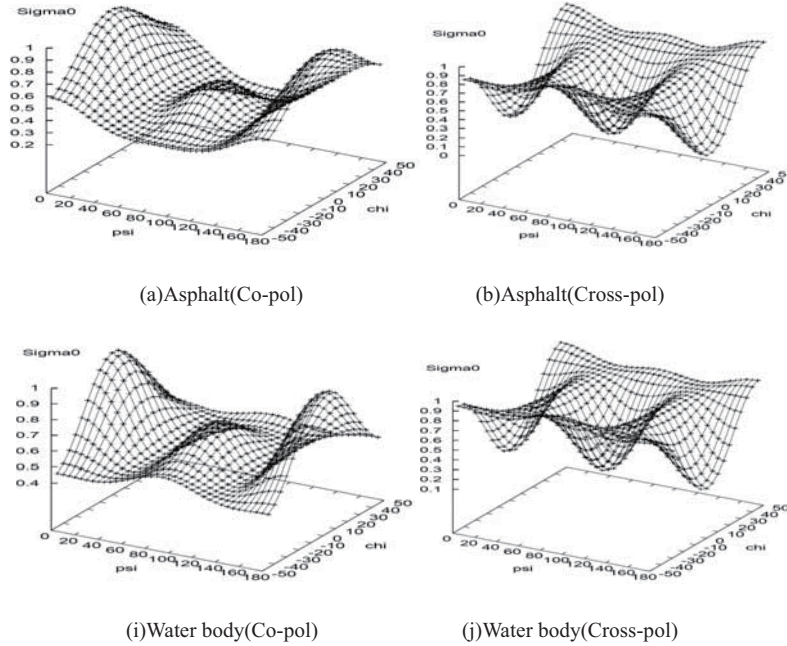


Figure A1 Polarization signatures of asphalt and water body

(2) Another example of classification performance for sea ice classification

The PI-SAR (Polarimetric and Interferometric SAR) data of the offshore of the lake Saroma in Hokkaido, Japan which was acquired by CRL (Communication Research Laboratory) and NASDA (National Space Development Agency of Japan) on 23 Feb. 1999. From the data of the SSC which consists of 5km x 5km of the area was extracted as is shown in Figure.A2(a): HH, (b): HV, (c): VV. Meanwhile, Figure.A2 (d) shows the maximum curvature of the trajectory of the cross polarized polarization signature projected in eigen space. The images are 8 look processed HH, HV and VV polarization of data which consists of 1000 x 987 pixels. In the figure, blue, white, black and green boxes show the training areas for the designated classes of TI (Nilas type of sea ice), SF, RF, and OW. Ice thickness of the TI is below 10 cm, and is covered with no snow on the surface while the SF ice thickness ranges from 10 to 70 cm and is covered with below 15 cm of snow. On the other hand, the RF ice thickness and snow cover situations are the same as SF while surface roughness is greater than SF due to the fact that the RF is formed surface roughness during the drift. The definitions of the classes are totally identical to the previous research work [19]. The definition is confirmed with Landsat-5 TM image which was acquired on 24 Feb. 1999 and is confirmed with visual perception from the shipment observation by one of the authors at that time.

The typical examples of the polarization signature and the trajectory for just four training samples for four classes are shown in Figure.A3. In the figure, (a) is the cross polarized data while (b) is the co-polarized data. The difference of the trajectory among the classes for the cross polarized data is greater than that for co-polarized data. The variance of the characteristics for both polarization signature and the trajectory for RF is the highest followed by TI, SF and OW. Also the maximum curvature of the trajectory of OW is the greatest followed by TI, SF and RF.

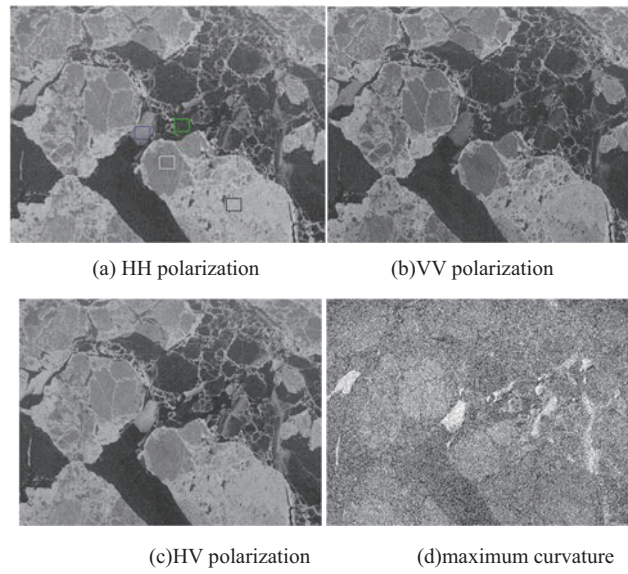
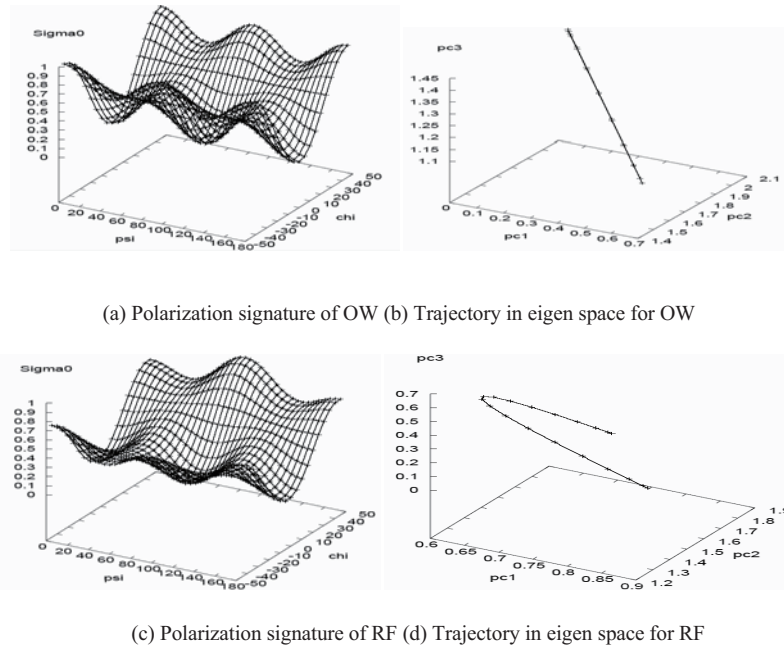
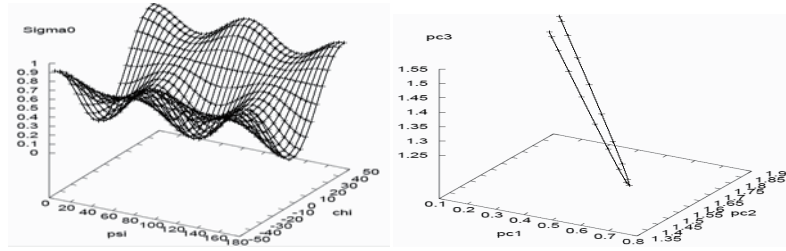
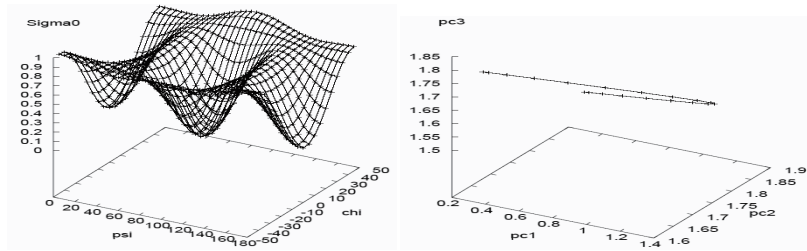


Figure.A2 Extracted PI-SAR imagery data of intensive study area of offshore of the lake Saroma in the sea of Okhotsk in Hokkaido vicinity in Japan acquired on 23 Feb. 1999.





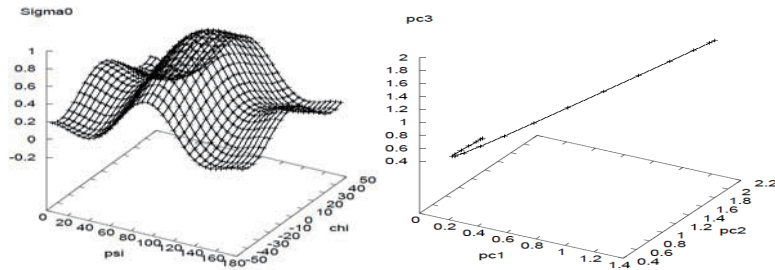
(e) Polarization signature of SF (f) Trajectory in eigen space for SF



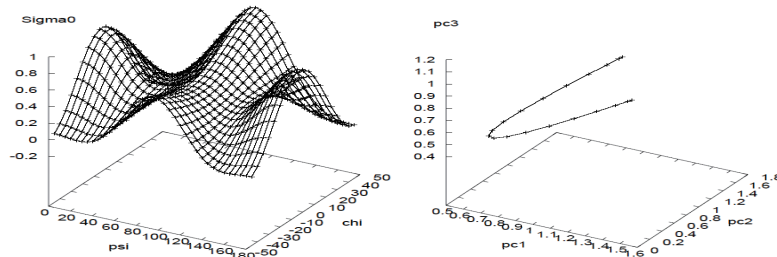
(g) Polarization signature of TI (h) Trajectory in eigen space for TI

Figure.A3(a) The typical cross polarization signature and the trajectory in eigen space converted from the 45 dimensional data with the quantization step of 5 degrees of σ_0 for polarization angle from HH to VV through HV for four classes, open water (OW), rough first year ice (RF), smooth first year ice (SF) and thin ice (TI) where PC1,2,3 denote the first to third principal components.

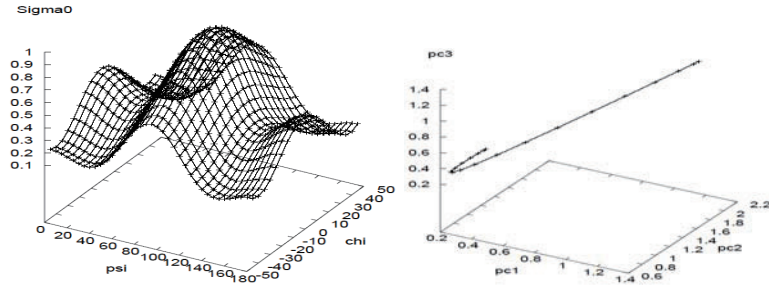
The confusion matrices of the proposed and the existing methods are shown in Table A1. The classes between SF and OW are most difficult to discriminate. It is followed by the classes between SF and RF and the classes between OW and SF.



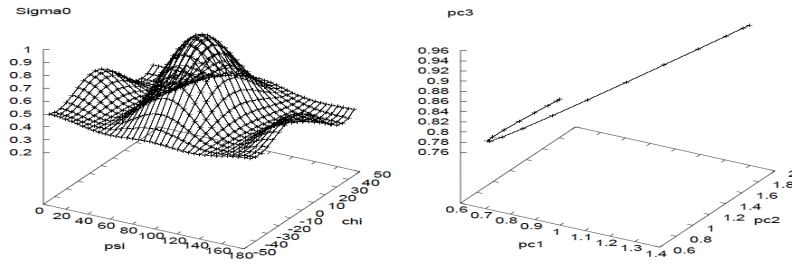
(a) Polarization signature of OW (b) Trajectory in eigen space for OW



(c) Polarization signature of RF (d) Trajectory in eigen space for RF



(e) Polarization signature of SF (f) Trajectory in eigen space for SF



(g) Polarization signature of TI (h) Trajectory in eigen space for TI

Figure.A3(b) The typical co-polarization signature and the trajectory in eigen space converted from the 45 dimensional data with the quantization step of 5 degree of σ_0 for polarization angle from HH to VV through HV for four classes, open water (OW), rough first year ice (RF), smooth first year ice (SF) and thin ice (TI).

As shown in Figure.A2, RF is easy to discriminate with the other classes even if only HH component used for the classification. Also, OW is easy to discriminate with the other classes. The correlation matrix among the classes shows that the difficult minimal pairs in the classification. In terms of scattering characteristics for OW, TI and SF are resembled so that it is not so easy to discriminate each other. It is clear that TI can be discriminated by using the maximum curvature as is indicated in Figure.A2 (d). This implies that the third principal component of TI differs from the other. As is shown in Figure.A3, the trajectory of OW is totally different from the other classes. Also the trajectory of RF shows the smallest maximum curvature and differs from the others classes so that the confusion probability between RF and the others is not so good. Consequently, it is found that Percent Correct Classification: PCC of the proposed method is superior to the maximum likelihood classification without the maximum curvature, the existing methods. Table A1 shows the confidence intervals: CI at the 95% confidence level for the correct classification probability.

$$CI = PCC \pm 1.96 \sqrt{\frac{\sigma^2}{n}} \quad (A1)$$

where σ denotes standard deviation of PCC and n denotes the number of training samples (approximately 1000 samples). Also the kappa coefficient k (a measure of pair-wise agreement) is evaluated.

TABLE A1 THE CONFUSION MATRICES OF (1) THE PROPOSED METHOD BASED ON MAXIMUM LIKELIHOOD CLASSIFICATION (MLH) USING THREE RECEIVED POWER, HH, HV AND VV AS WELL AS THE MAXIMUM CURVATURE OF THE TRAJECTORY OF $\Phi\sigma^0$ IN THE SPACE OF ELLIPTICITY AND ORIENTATION ANGLES, (2) USING JUST THREE RECEIVED POWER, (3) USING THREE RECEIVED POWER + H + A + A COMPONENTS AND (4) USING THREE RECEIVED POWER + ODD + EVEN + DIFFUSE COMPONENTS (C.I. DENOTES CONFIDENCE INTERVAL AT THE 95% OF CONFIDENCE LEVEL AND K DENOTES KAPPA COEFFICIENT).

(1) PROPOSED METHOD (PCC=95.7), $K=0.899$ (2) USING HH + HV + VV (PCC=93.63), $K=0.886$

	OW	RF	SF	TI	C.I.		OW	RF	SF	TI	C.I.
OW	96.4	1.2	2.4	0.0	0.32	OW	91.2	1.3	7.5	0.0	0.28
RF	2.3	92.2	5.5	0.0	0.41	RF	2.5	91.2	6.3	0.0	0.39
SF	2.3	0.8	96.5	0.4	0.55	SF	3.4	0.7	94.3	1.6	0.51
TI	0.0	0.0	2.3	97.7	1.02	TI	0.0	0.0	2.2	97.8	1.01

(3) USING HH+HV+VV+H+A+A(PCC=69.95), $K=0.591$ (4) USING HH+HV+VV+ODD+EVEN+DIFFUSE (PCC=93.38), $K=0.885$

	OW	RF	SF	TI	C.I.		OW	RF	SF	TI	C.I.
OW	88.7	3.9	7.4	0.0	0.25	OW	93.4	0.8	5.8	0.0	0.30
RF	3.5	91.8	4.7	0.0	0.40	RF	3.1	92.2	4.7	0.0	0.41
SF	21.1	2.3	76.6	0.4	0.45	SF	6.6	0.8	92.2	0.4	0.49
TI	29.3	0.4	47.7	22.7	2.06	TI	0.8	0.0	3.5	95.7	0.99

Although H, A and α , in general, does work for classification, the distribution of the data of TI and SF are overlapped in the H, A and α in the feature space. Therefore, the confusion probability between smooth first year ice and thin ice is relatively large followed by the classes between open water. Meanwhile, odd, even and diffuse scattering components work for classification, in particular, the discrimination of RF with the other classes due to the fact that the scattering mechanism of RF is different from the other classes. However, its effectiveness is weak so that the classification performance is not so good accordingly.

It is found that the proposed method shows around 2.2-4.0% of improvement of classification performance for the sea ice (Okhotsk) classifications in comparison to the classification with just three received power of HH, HV, VV polarization signals. Furthermore, it is also found that the classification performance for the Tsukuba data is improved by 14.3% for the proposed method. Thus, it is concluded that the newly proposed feature of the maximum curvature of the trajectory of polarization signature projected in eigen space through principal component analysis is effective for classifications.

Acknowledgements

The author would like to thank Dr. J. Wang (Former graduate school student of Saga University) for her contributions to this research works.

References

1. Zebker, H.A., vanZyl, J.J. Imaging radar polarimetry: A review. Proc. IEEE 79, 1583–1606, 1991.
2. Ulaby, F.T., Elachi, C. Radar Polarimetry for Geoscience Applications. Artech House Inc., Dedham, 1990.

3. Mott, M. Antennas for Radar Communications – A Polarimetric Approach. John Wiley, New York, 1992.
4. Henderson, F.M., Lewis, A.J. Principles and applications of imaging radar, in: Manual of Remote Sensing, third ed. John Wiley, New York, 1998.
5. Krogager, E. Analysis of the absolute and relative phase conditions of transforming matrices for the sinclair and covariance matrices in radar polarimetry. J. Opt. Soc. Am. E72B, 1A, 1993.
6. Dong, Y. Imaging radar polarimeter from wave synthesis. J. Geophys. Res. 92 (81), 638–701, 1987.
7. Zebker, H.A., et al., Imaging radar polarimeter from wave synthesis, J. Geophysical Research, Vol.92, No.81, pp.638-701, 1987.
8. Krogager, E., Czyz, Z.H. Properties of the sphere, di-plane, helix decomposition, in: Proceedings, Third International Workshop on Radar Polarimetry, vol. 1, pp. 106–114, 1995.
9. Cloude, S.R., Pottier, E. A review of target decomposition theorems in radar polarimetry. IEEE Trans. Geosci. Remote Sensing 34 (2), 498–518, 1995.
10. Pottier, E., Lee, J.S. Unsupervised classification scheme of PolSAR images based on the complex Wishart distribution and H/A/a polarimetric decomposition theorems, Proceedings of the Third European Conference on Synthetic Aperture Radar: EUSAR, 2000.
11. Scheuchl, B., Hajnsek, I., Cumming, I.G., et al. Sea ice classification using multi-frequency polarimetric SAR data, in: Proceedings of the IGARSS'02, 2002.
12. vanZyl, J.J., Zebker, H.A., Elachi, C. Imaging radar polarization signatures, theory and observation. Radio Sci. 22 (4), 529–543, 1987.]
13. Kohei Arai and J.Wang, Polarimetric SAR image classification with maximum curvature of the trajectory in eigen space domain on the polarization signature, Advances in Space Research, 39, 1, 149-154, 2007.
14. Kohei Arai, Polarimetric SAR image classification with high frequency component derived from wavelet multi resolution analysis: MRA, International Journal of Advanced Computer Science and Applications, 2, 9, 37-42, 2011
15. Kohei Arai Comparative study of polarimetric SAR classification methods including proposed method with maximum curvature of trajectory of backscattering cross section in ellipticity and orientation angle space, International Journal of Research and Reviews on Computer Science, 2, 4, 1005-1009, 2011.
16. Kohei.Arai, Fundamental Theory on Wavelet Analysis, Morikita-Shuppan Publishing Co. Ltd., 2000.
17. Kohei.Arai, Leland Jameson, Earth Observation Satellite Data Utilization Methods by Means of Wavelet Analysis, Morikita-Shuppan Publishing Co. Ltd., 2001.
18. Kohei.Arai, Lecture Note on Wavelet Analysis, Kindai-Kagaku-Shuppan Publishing Co. Ltd., 2006.
19. Matsuoka, T. et al., CRL/NASDA airborne SAR (PI-SAR) observations of sea ice in the sea of Okhotsk, Annals of Glaciology, 33, 115-119, 2001

コンクリート建造物に対する超音波 CT 開発と窓関数の応用

滝口 孝志 *

* 防衛大学校数学教育室

概要. コンクリート建造物に対する tomographic な非破壊検査技術は未開発であるというのが現状であるが、近年、三田紀行 (職業能力開発総合大学校) と講演者の協働により、コンクリート建造物に対する acoustic CT 開発のアイデアが提案されている。本講演では、コンクリート建造物に対する acoustic CT 開発のアイデアを紹介し、この問題に対する窓 Radon 変換の応用に関して議論する。

Development of ultrasonic CT for concrete structures and application of window functions

Takashi Takiguchi*

*National Defense Academy of Japan

Abstract. There having been developed no tomographic non-destructive testing technique for concrete structures for the time being, N. Mita and the speaker proposed an idea to develop an acoustic CT for concrete structures. In this talk, we introduce their idea and discuss application of the windowed Radon transform for development of an acoustic CT for concrete structures.

1. はじめに

コンクリート建造物に対する tomographic な非破壊検査技術は未開発であるというのが現状であるが、近年、様々な手法が提案され始めている。三田紀行 (職業能力開発総合大学校) と講演者は、[2,3] においてコンクリート建造物に対する超音波 CT 開発のアイデアを提案している。本講演では、コンクリート建造物に対する超音波 CT 開発のアイデアを紹介し、この問題に対する窓 Radon 変換の応用に関して議論する。

2. コンクリートの基礎知識

本節では、本講演で必要となるコンクリートの基礎知識について簡単に説明する。

定義 2.1 (コンクリート) コンクリートは、セメント (cement: C)、水 (W) 砂 (fine aggregate or sand: S), 砂利 (coarse aggregate or gravel: G) の混合物である。

- 注意 2.1 (i) セメント + 水 → セメントペースト (cement paste)
 (ii) セメント + 水 + 砂 (セメントペースト + 砂) → モルタル (mortar)
 (iii) コンクリート (concrete) は “モルタル + 砂利”
 と理解できる。

本節では、以下の主張に基づき、コンクリートとは何かを説明する。

主張 2.1 *The concrete materials are artificial megaliths.*

主張 2.1 は [2,3] に於いて主張されたものであるが、以下の議論に基づいている。例えば, Valley Temple (BC2500?) や Parthenon 神殿 (BC447-432) は石造建造物であり, 築後 2000 年以上経った現在でも, その構造を保っている。ところが, ローマの Colosseum (AD70-80) の外壁は石造であるが, 内壁は, 煉瓦の破片・石・砂・灰等で埋められている。これは, ローマでは十分な巨石が入手できなかったためであるが, Colosseum の内壁に用いられた素材 (Roman or primitive concrete) は 2000 年近く巨石の代替物としての役割を果たしてきたといえる。

コンクリートが巨石の代替物になるには,

- (a) Concrete is not tough to tensile strength.
- (b) Concrete easily gets cracks in and on itself.

等の問題点を改善させなければならないが,

- (α) Concrete is easily made and shaped in any form because of its fluidity before it gets hard.
- (β) The cost of concrete is very cheap.

等, 巨石より優れた点も多い。欠点 (a) を補うために開発されたのが鉄筋コンクリート (RC) である。

- 鉄とコンクリートとの熱膨張係数はほぼ同じである
- セメントのアルカリ性が鉄筋を錆から守る

等の理由から, 鉄筋コンクリートが開発された当時は, 鉄筋コンクリートは, “奇跡の素材”, “永久素材” 等と呼ばれていたが, 後にそれらが幻影であったことがわかった。現在は, RC 建造物の長寿命化とその維持管理がインフラの維持管理・再整備において重要な課題の一つとなっている。

3. コンクリート建造物に対する CT

前節で述べた通り, RC 建造物の長寿命化とその維持管理は非常に重要な問題であるが, コンクリート建造物に対する tomographic な非破壊検査法は確立されていないというのが現状である。近年, 様々な主張が提案されているが, 三田と講演者は [2,3] において, 超音波や電磁パルス音響を用いた acoustic CT scan の開発を提案している。本節では, [2,3] で提案されたアイデアを紹介する。

彼らは, まず, 実験を通じて, 以下の性質を確認した。

性質 3.1

- *We can conclude that for the test pieces of the length less than 1200mm, there is no decay of the acoustic velocity from the viewpoint of its first arriving time.*
- *The first arrival wave of the ultrasonic one takes the fastest route in the cement paste, the mortar and the concrete.*

これらの性質を用いて, コンクリート建造物に対する acoustic CT 開発のためには次の問題を解決すれば良いことを主張している。

問題 3.1 (Problem for non-destructive inspection for concrete structure) *Let $\Omega \subset \mathbb{R}^3$ be a domain and $f(x)$ be the propagation speed of the ultrasonic wave at the point $x \in \Omega$. For $\alpha, \beta \in \partial\Omega$, we denote by $\gamma_{\alpha, \beta}$ a route from α to β contained in Ω . In this case, reconstruct $f(x)$ ($x \in \Omega$) out of the data*

$$(3.1) \quad \min_{\gamma_{\alpha, \beta}} \int_{\gamma_{\alpha, \beta}} 1/f(x) d\gamma,$$

for $\forall \alpha, \beta \in \partial\Omega$.

性質 3.1 にあるようにこの問題は長さが 1.2m 以下の建造物を対象にしているが, 鉄筋高層建造物の柱, 高速道路橋梁におけるかぶりコンクリートの異常調査等, 多くの問題への応用が検討されている。詳しくは [3] を参照されたい。

4. 電磁パルス音響装置

前節で紹介した acoustic CT では, 超音波装置も電磁パルス装置も音源と受信装置の組み合わせで用いられており, 原理的に両者の違いは無いが, 既存の装置を用いた場合, 電磁パルス装置の方が観測の精度が低い。勿論, 観測精度は装置の精度に依存するため, 一概には結論付けは出来ないが, 電磁パルス装置による観測精度の低さの原因は, パルス音発生 の原理に依るところが少なくないように感じられる。

電磁パルスによる音波発生は以下のような原理に基づいている。

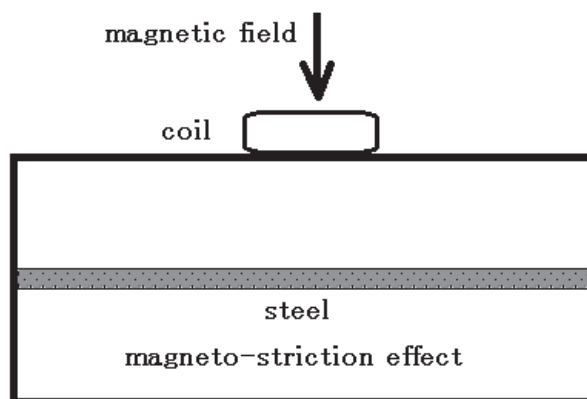


Fig. 1. 電磁パルス音発生の原理

図 1 において, 図の鉛直方法に磁場をかけ, コイルにパルス電流を流すことにより, 装置内の金属 (鉄) が磁歪効果 (magneto-striction effect) により超高音を発生する。従って, 電磁パルス音は超音波であり, その周波数を調整する技術も進んでいるが, 金属のどの部分から音を発しているかは厳密には調べられていない。勿論, 電磁パルス音響装置内の金属は微小であるため, 大きな誤差は生じていないと思われるが, もし, 観測毎に音源の場所が変わっていれば, 観測地の精度に大きな影響を与えているはずである。

本講演では, 鉄筋コンクリート若しくは鉄筋モルタルに対して, 直接電磁パルスをぶつけることにより, 鉄筋のどこから音が出ているかを調べる問題を考える。

問題 4.1 鉄筋コンクリート, 若しくは鉄筋モルタル, に対する電磁パルス音響に於いて, 音源点を全て特定せよ。

この問題が解決すれば, 内部注入を施さない RC 建造物に対する音響 PET (Positron Emission Tomography) への応用が可能になる。本講演では, 以下の 3 点を紹介することを主目標とする。

- 問題 4.1 が次節で紹介する切断窓 X 線変換と密接な関係を持つことを示すこと。
- 幾つかの想定に対して問題 4.1 が解決可能であることを示し, その解を与えること。
- 問題 4.1 の解析結果を用いて, 内部注入を施さない RC 建造物に対する音響 PET 開発が可能であること。

5. 窓 Radon 変換

本節では、前節の最後に提起した問題 4.1 に関連する積分変換を紹介する。本節の主目標は切断窓 X 線変換の定義を与えることであるが、まずは、その基になっている、窓 Radon 変換 (窓 X 線変換) の定義を復習する。窓 Radon 変換 (窓 X 線変換) は Radon 変換 (X 線変換) に窓関数を掛けた積分変換であるが、その定義は以下で与えられる。

定義 5.1 (窓 Radon 変換) *Let $1 \leq d \leq n$. The d dimensional Windowed Radon Transform $R_h f(x, A)$ of a function f defined on $\mathbb{R}^n$ is defined by*

$$(5.1) \quad R_h f(x, A) \equiv \int_{\mathbb{R}^d} \overline{h(t)} f(x + At) dt,$$

where h is a function defined on $\mathbb{R}^d$ and $A \in \{A = \nu R J \mid J : \mathbb{R}^d \mapsto \mathbb{R}^n \text{ (inclusion)}, R \in SO(n)/SO(n-d), \nu > 0\}$. $\bar{h}$ means the complex conjugate of h . We write $R_h f(x, A) = f_h(x, A)$.

定義 5.2 (窓 X 線変換) *For a function f defined on $\mathbb{R}^n$ we call*

$$(5.2) \quad X_h f(x, \xi) \equiv \int_{\mathbb{R}} \overline{h(t)} f(x + t\xi) dt,$$

the Windowed X-ray Transform (WXT) with window h , where h is a function defined on $\mathbb{R}$ and $\xi \in \mathbb{R}^n$. We adopt the notation $\xi = \nu\omega$, $\nu > 0$, $\omega \in S^{n-1}$.

これらの変換に対する性質などは、[4] を参照されたい。

詳しくは講演にて説明するが、問題 4.1 において、最も単純な場合には、以下の逆問題を解決すれば良いことがいえる。

問題 5.1 $\mathbb{R}^2$ 上で定義された関数 $f(x)$, $x \in \mathbb{R}^2$ が既知である時、観測値

$$(5.3) \quad \int_a^\infty f(y + t\theta) dt$$

から a を復元せよ。但し、 $\theta \in S^1$, $y \perp \theta$ とする。

(5.3) は切断 X 線変換と呼ばれる変換であるが、問題 4.1 に関しては、この変換が切断窓 X 線変換に対応する事がわかる。詳しくは講演にて説明する。

6. おわりに

本講演に於いて、電磁パルス音響における音源の特定に関する問題を提起したが、この問題は切断窓 X 線変換における切断点を復元する問題 5.1 と深い関係を持っている。問

題 4.1 の研究を発展させると、コンクリート建造物に対する音響 PET の開発が可能になる。今後、実験と理論の両者の立場からこれらの問題の研究を進め、コンクリート建造物の非破壊検査法のさらなる発展に繋げたい。

謝辞 本研究は、JSPS 科学研究費 (C)26400184 の助成を受けています。

参考文献

- [1] Mita, N., Horie, Y. and Matsudome, S. : *Fundamental study on estimation of compressive strength of cement mortar by pulsed electromagnetic force acoustic method*, Architectural Institute of Japan **76** (662) (2011) , pp. 721–727.
- [2] Mita, N. and Takiguchi, T. : *Basic propertiers of concrete and its non destructive testing*, MI Research **2**, “Collaboration between theory and practice in inverse problems”, (2015), pp. 117-137.
- [3] — — : *Development of ultrasonic tomography for concrete structures*, MI Research **5**, “Mathematical backgrounds and future progress of practical inverse problems”, (2016), pp. 205-233.
- [4] Takiguchi, T. : *On invertibility of the windowed Radon transform*, J. Math. Sci., Univ. Tokyo, **2** (1995) pp. 621-636.

滝口 孝志 (防衛大学校数学教育室)

〒239-8686 神奈川県横須賀市走水 1-10-20

E-mail: takashi@nda.ac.jp



国立大学法人

大阪教育大学



統計数理研究所

数学協働プログラム

ウェーブレット理論と工学への応用

主催 大阪教育大学・統計数理研究所

平成二十八年十一月