

平成 25 (2013) 年度援助 研究課題報告書

提出:2014年 3月 1日

日本医学物理学会 御中

1. 研究課題名

日本語	逐次近似型画像再構成法における視覚的印象を反映した新しい画質評価法の開発
英語	Iterative tomographic reconstruction: quantitative assessment of perceptual image quality with a novel statistical metric

2. 研究代表者

(ふりがな)	やまだ さちこ	JSMP 会員番号	001712
氏名	山田 幸子		
所属機関	大阪大学		
部署	医学部附属病院 医療技術部		
役職	副主任医療技術職員	E-mail	yamada@hp-rad.med.osaka-u.ac.jp
所在地	〒565-0871 大阪府吹田市山田丘2-15		
電話	06-6879-6812	FAX	06-6879-6814

3. 研究組織

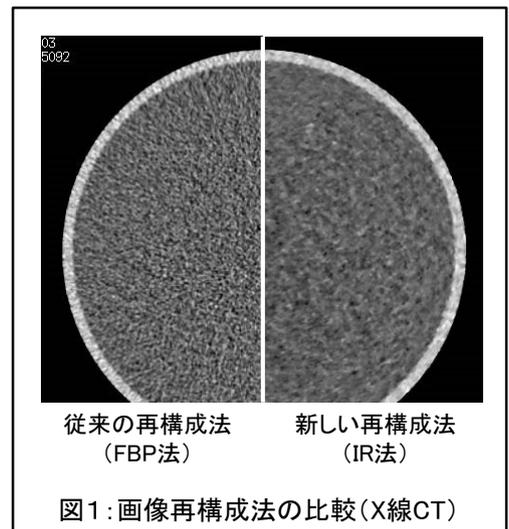
氏名	所属機関
上口 貴志	情報通信研究機構
尾方 俊至	神戸低侵襲がん医療センター
荻原 良太	大阪大学

## 4. 研究目的、背景および当初の目標

### 4. 1. 研究背景および目的

X線CT等の医用イメージングにおいては、画像診断の目的を達成できる画質をできるだけ低線量で得る必要がある。ゆえに画像診断の目的に応じた撮影法や画像再構成法の最適化が必要であり、そのためには客観的、定量的な画質評価が不可欠となる。従来これは、鮮鋭度やノイズ、コントラストといった古典的な画質因子で評価されてきた。しかし近年、たとえばX線CTにおいては、大幅な被ばく低減を目的に新しい画像再構成法が導入され、その画質をめぐる激しい議論が起きている。

従来のX線CTでは、フィルタ補正逆投影(filtered back-projection; FBP)法、またはそれと数学的に等価なアルゴリズムが画像再構成の標準的方法であった。これに対し、近年では逐次近似型画像再構成(iterative reconstruction; IR)法<sup>(注1)</sup>が広く臨床応用され、FBP法に対して鮮鋭度やコントラストを損なうことなくノイズの大幅な低減が可能であるとの研究報告が多くなされている。しかし、IR法で得られた画像は従来のFBP法で得られたものとは異質であることも指摘されており、従来の画質評価結果と視覚評価結果のミスマッチが問題となっている。たとえば、図1は2種類の再構成法によるファントム画像の例を示すが、FBP法による再構成画像(左)に比べてIR法による画像(右)は明らかに異なる視覚的印象を与える。しかし、それを客観的に表す指標は存在せず、臨床の場では「(FBP画像に比べてIR画像は)違和感を覚える」といったあいまいな表現で形容されているにすぎない。公正な画質評価を行い、撮像法や画像再構成法を最適化するためには、視覚的印象の定量化が不可欠である。本研究の目的は、「違和感を覚える」と形容されるIR画像の主観的印象を定量的に扱うための方法を開発することである。



(注1)IR法はFBP法に比べて画像再構成に多大な時間を要するため、データ量が多いX線CTではまず、FBP法で画像を再構成し、それにノイズ低減のためのIR処理を加える形で実装されることが多い。本研究では、このような画像再構成法も広くIR法と称する。臨床用装置においては、FBP画像にIR処理をどの程度加えるかをパラメータとして与えられることが多く、たとえばSiemens社のSAFIRE(Sinogram Affirmed Iterative Reconstruction)法では5段階、GE社のASiR(Adaptive Statistical Iterative Reconstruction)法では10段階のノイズ低減レベルを設定可能である。以下、このパラメータをIRレベルという。高いIRレベルを設定するほど、すなわちノイズ低減を強く図るほど、得られる画像はFBP画像に比べて違和感の強いものとなる。

## 4. 2. 当初の目標

われわれは、FBP画像に対するIR画像の違和感が、画像上のテクスチャの変化として定量できないかと考えた。テクスチャとは、元来、布地の織り目を指す語であり、画像解析の分野では画像の特徴量として広く用いられている。テクスチャ特徴量にはさまざまなものが知られているが、IR レベルを変化させたときにそれに相関をもって量が増えるようなテクスチャ特徴量を探索することで、視覚的印象を定量的に扱える指標が見出せるのではないかと考えた。そこで、さまざまなテクスチャ特徴量のなかから本研究の目的を達成しうる候補を絞り込み、さらに視覚的評価を経て新しい画質評価のための指標として確立する計画を立てた。

## 5. 研究成果概要:得られた成果

### 5. 1. 昨年度の研究成果の概要(昨年度の研究成果報告書より要点のみを抜粋)

まず、本研究の背景となった、古典的画質評価法の問題点を確認するため、IR レベルとノイズ、解像度、およびコントラストの関係を評価した。その結果、IRレベルを高めるにしたがって、(1)ノイズは単調かつ一定の割合で減少、(2)解像度は不変、(3)コントラストも不変、という結果を得た。以上の結果から、「IR 法は、解像度やコントラストに影響を与えることなくノイズを低減でき、最大の IR レベルを使用することでその効果は最大となる」旨の結論が誘導されることを確認した。また、IR レベルを大きくするに従い、FBP 画像との視覚的印象の相違が明瞭となり、不自然な画像になることが確認された。したがって、上で述べた古典的な画質評価結果は、IR 法での撮像、画像再構成法の最適化には不完全(または不向き)であることが確認された。

つぎに、当初の目標であるテクスチャ特徴量探索に向けて検討を開始した。しかしその序盤、ノイズパワースペクトルを用いた IR 画像の画質評価の試みが他の研究チームより報告されるようになった。ノイズパワースペクトルはテクスチャ特徴量の一つと関係が深く、われわれも重要視していたものであるが、他の研究チームがすでに検討結果を提出し始めたことから、われわれは研究方針を修正し、まったく異なる発想にもとづく画質評価法を検討することとした。

われわれが採用した方針は、基準となるFBP画像に対するIR画像の「非類似性」の定量化である。すなわち、IR画像が違和感を与えるというのは、言い換えるとFBP画像と「似ていない」からであって、その非類似性の程度を定量できれば視覚的印象の違いを定量できるのではないかと考えた。このような比較法の一つとして、平均2乗誤差(MSE; mean square error)が知られている。これは基準画像と比較したい画像の平均2乗誤差を評価するものであるが、視覚的印象とは一致しないことも知られている。そこで、医用分野に限らず、より広く一般に文献を調査したところ、画像の輝度、コントラスト、および構造という3つの要素をそれぞれ比較し、統合することで、視覚的印象に沿った類似度の評価が可能であるとの情報を得た。そこで、FBP画像 $\mathbf{x}$ とIR画像 $\mathbf{y}$ に対してこの3要素により構成される非類似性評価関数 $M(\mathbf{x}, \mathbf{y})$ を定め、IRレベルを変化させたときの $M(\mathbf{x}, \mathbf{y})$ の振る舞いを評価したところ、IRレベルを大きくするにしたがってFBP画像との非類似性が高くなるという結果を得ることができ、これは

視覚的印象を反映した新しい画質評価指標になりうると結論づけた。

## 5. 2. 昨年度研究成果の問題点と貴学会からのご意見

非類似性評価関数  $M(\mathbf{x}, \mathbf{y})$  の出力は IR レベルに強く依存することから、FBP 画像と IR 画像の違いを反映していることに疑う余地はなかった。しかし、IR レベルを大きくすることで画像ノイズが低減されることから、 $M(\mathbf{x}, \mathbf{y})$  の評価している非類似性とは単に IR レベルに依存するノイズレベルの変動を見ているのではないかという懸念があった。そこで IR レベルを変化させたときの  $M(\mathbf{x}, \mathbf{y})$  とノイズレベルの挙動を調べたところ、両者には大きな乖離があり、単にノイズレベルの変動だけを評価しているのではないことは確認できたが、それが視覚的印象としての「違和感」を本当に反映しているのかという問題には依然として疑問が残った。また、貴学会理事会より、画像のテクスチャが低コントラスト構造の描出に影響を与えることから、ROC 解析など、低コントラストファントムを用いた視覚評価実験を検討すべきである旨のコメントをいただいた。以上を踏まえて今年度の研究計画を立案し、次項で述べる成果を得た。

## 5. 3. 今年度の研究成果

### 5. 3. 1. 評価用試料作成法の再検討

ファントムには、低コントラスト分解能評価用モジュールを含む市販品を採用した。また、FBP 画像と IR 画像にノイズレベルの違いがあると、それが評価関数の出力を左右することから、ノイズレベルの等しい FBP 画像・IR 画像対の作成を試みた。まず、ファントムをさまざまな線量にて撮影し、FBP 法および IR 法での画像再構成を行った。IR 法では、すべての IR レベルで再構成を行った。こうして得られた画像群に対しノイズ測定を行った。ノイズレベルは、均一な関心領域での CT 値の標準偏差として求めた。つぎに、FBP 画像と IR 画像のすべての組み合わせのなかから、ノイズレベルの等しい画像対を選出し、それらを評価対象試料とした。なお、ファントムには低コントラストモジュールのほか、空間分解能評価用高コントラストモジュールやコントラスト評価等の多目的モジュールも含まれ、いずれのモジュールにおいてもノイズレベルの等しい FBP 画像・IR 画像対を作成した。

### 5. 3. 2. 非類似性評価関数と視覚的評価との対比

ノイズレベルの等しい FBP・IR 画像対に対して、昨年度に構築した非類似性評価関数の振る舞いを評価した。また、低コントラスト検出能について視覚評価を行い、評価関数の出力と対比した。その結果、IR レベルが高くなるほど、FBP 画像に対する非類似度も高いと判定されるものの、その評価結果は必ずしも視覚的な印象とは一致しない傾向が認められた。その原因を調査したところ、以下の点に問題があることが明らかとなった。

(1) CT 画像は -1000HU から +3000HU 超まで幅広いダイナミックレンジをもっており、IR 法によるテクスチャの変動は水濃度に近い構造物、空中、高吸収な構造物のいずれにも出現し、評価関数はそれらすべての変化を捉える。一方、画像観察は適切なウィンドウレベル、ウィンドウ幅にて行われる。そして軟部組織など水濃度に近い構造物

を観察する際には空中は「黒」、骨など高濃度なものは「白」として表示され、空中や高濃度領域に発生するテクスチャの変動は視認できない。同様に、高吸収なものを観察する際にはウィンドウレベルを高く、ウィンドウ幅を広く設定するが、これによって水濃度に近い構造物でのテクスチャの変動はほとんど視認できなくなる。したがって、視覚的印象を定量化するうえでは、画像観察の目的に応じたウィンドウ設定を考慮に入れる必要があることが明らかとなった。

(2)ROC 解析を含めた低コントラスト構造物の視覚評価においては、おもに構造物の検出感度が評価される。しかし、IR 法で得られる違和感の強い画像は、「検出は容易だが明らかに画質が劣化している」ものが多く、検出感度の評価のみでは不十分であることがわかった。たとえば図2は円形低コントラスト構造物の描出を比較したもので、両者のノイズレベルは同等である。FBP 法、IR 法とも構造物の検出は容易であるが、IR 法ではその形状が明らかに円形とは異な

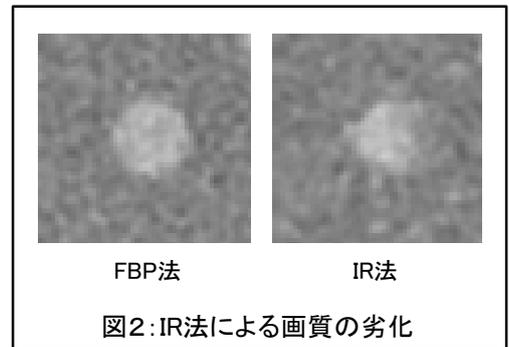


図2:IR法による画質の劣化

って認められる。ここでは比較的コントラストが強く径の大きなものを例示したが、より低コントラストなものや小径のものにも同様の傾向が認められた。もちろん、検出感度をROC 解析等で評価すること自体は有用であるが、本研究の目的に立ち返ると、感度に差がなくとも図2に示すような視覚的印象の違いを表現できる方法を開拓することが最優先される課題であり、視覚的評価の方法としては検出感度ではなく、より主観的な印象である「違和感」そのものを観察者に回答してもらうデザインを選択した。

### 5. 3. 3. 違和感の定量的評価法の再構築

FBP 画像とIR 画像の定量的な比較に際し、ウィンドウ設定の影響を加味するため、両 CT 画像に対するダイナミックレンジ狭小化処理を加えることとした。すなわち、画像観察の対象や目的に応じて最適なウィンドウレベル、ウィンドウ幅を決定し、その範囲を8ビットの正の画素値へと変換した(ウィンドウ外は0もしくは255の値となる)。つぎに、両画像の類似性に着目した評価関数 SSIM(structural similarity)を導入した。FBP 画像およびIR 画像をそれぞれ  $f(x, y)$ 、 $g(x, y)$  とすると、評価関数  $SSIM(f, g)$  は両画像の着目画素  $(x, y)$  近傍における輝度の類似性  $l(f, g)$ 、画素値の変化(コントラスト)の類似性  $c(f, g)$ 、構造の類似性  $s(f, g)$ 、正の係数  $\alpha$ 、 $\beta$ 、 $\gamma$  を用いて  $SSIM(f, g) = [l(f, g)]^\alpha [c(f, g)]^\beta [s(f, g)]^\gamma$  と表現される。ここで  $l(f, g)$ 、 $c(f, g)$  は感覚に関する基本法則であるウェーバーの法則に近似的に従う関数形をとり、また  $s(f, g)$  は近傍の輝度とコントラストで正規化した画素値分布の相関である。SSIM は、 $f(x, y)$  と  $g(x, y)$  とが完全に一致すれば最大値 1.0 をとり、完全に独立であれば 0.0 となる。そして着目画素  $(x, y)$  近傍における違和感 PID(perceptual image dissimilarity)の定量値を  $PID(f, g) = 1.0 - |SSIM(f, g)|$  と定義した。PID は座標  $(x, y)$  の関数でもあり、その空間分布を以下、違和感マップという。違和感マップに任意の関心領域を設定し、PID の平均値を求めれば、それが当該関心領域での違和感の定量値を与え、関心領域をファントム全体に設定すれば、当該ファントム全体を観察したときの違和感を与える。PID の妥当性は視覚的評価との対比を行うことで検証した(後述)。

#### 5. 3. 4. 平均2乗誤差(MSE)の評価

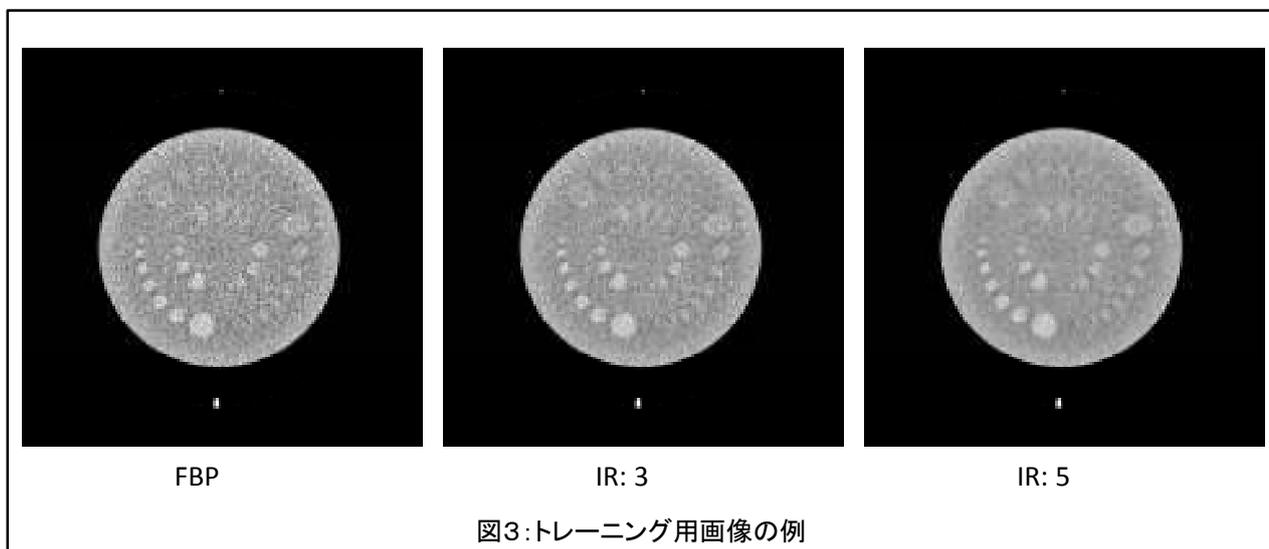
最近の研究動向として, FBP 画像に対する IR 画像の違和感を MSE で評価しようとする報告が見られる. そこで, MSE にて違和感を正しく評価しうるかを検討するため, FBP 法と IR 法の差分画像の各画素値を2乗した「2乗誤差 (SE;squared error) マップ」を求めた. SE マップを全画素について平均したものが MSE に一致する.

#### 5. 3. 5. 違和感の視覚的評価

X線CT画像の画質評価に十分な経験を有する観察者5名により, FBP 画像に対する IR 画像の違和感について視覚評価を実施した. ここで違和感とは, 臨床画像で問題とされるテクスチャの変化と定義し, IR レベルを変化させたときのテクスチャの変化を示した画像(図3に例示)を用いてトレーニングを行ったのち, 本実験を開始した. 提示画像は5. 3. 1. で述べた FBP 画像・IR 画像の組み合わせ60対とし, 各対において FBP 画像と IR 画像のノイズレベルは等しい. 観察者は, FBP 画像に対して IR 画像にどの程度違和感を覚えたかを 5 段階スケール(違和感スコア)で回答し, ウィンドウ設定は, 画質評価用ファントムに含まれる3種のモジュールに対して以下の条件とした.

- (1) 高コントラスト評価用モジュール: WW=800, WL=400 および WW=100, WL=100 の2種類
- (2) 多目的評価用モジュール: WW=200, WL=100, および WW=100, WL=100 の2種類
- (3) 低コントラスト評価用モジュール: WW=40, WL=50, および WW=100, WL=50 の2種類.

ここで, WW, WL はウィンドウ幅, ウィンドウレベルであり, 単位はいずれも HU である.



#### 5. 3. 6. 結果

図4(次ページ)に, 本研究で構築した FBP 画像に対する IR 画像の違和感定量値 PID と視覚評価での違和感スコアとの対比を示す. また, 図5(次ページ)に FBP 画像, IR 画像間の平均2乗誤差 MSE と違和感スコアとの対比を示す. いずれも, 観察者の回答した5段階の違和感スコアに対する PID 値, MSE 値をプロットしたもので, 5名の観察者間のばらつき(標準偏差)をエラーバーとして表示している. PID, MSE とも, 数値が大きいほど「違和感

が強い」と評価されたことになる。

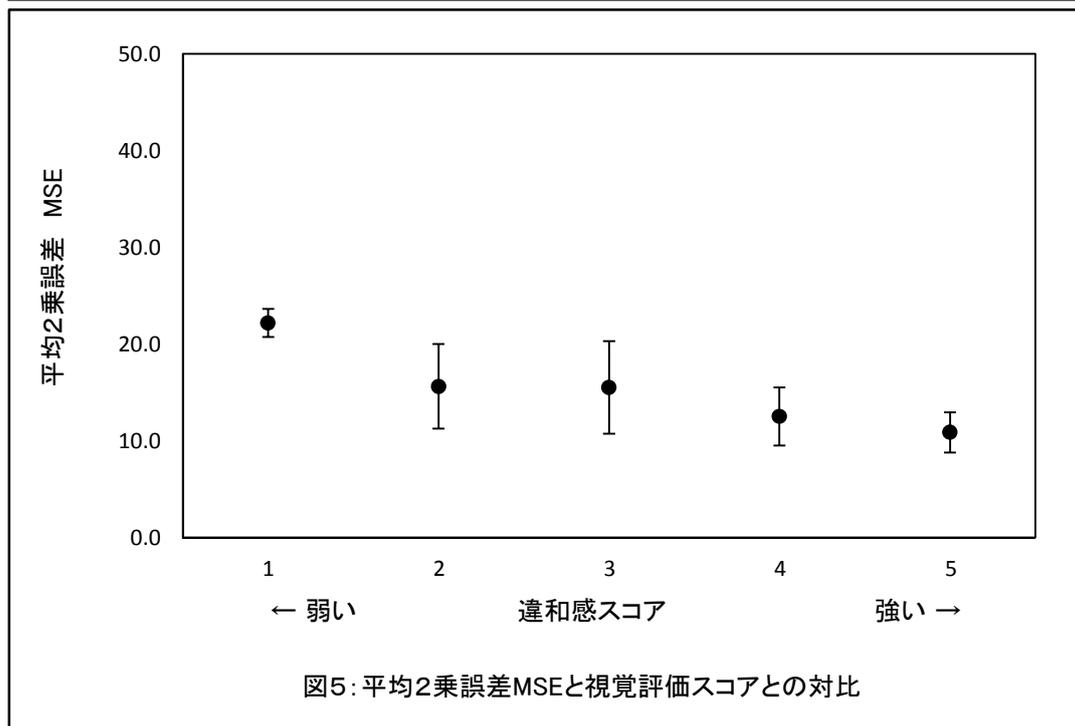
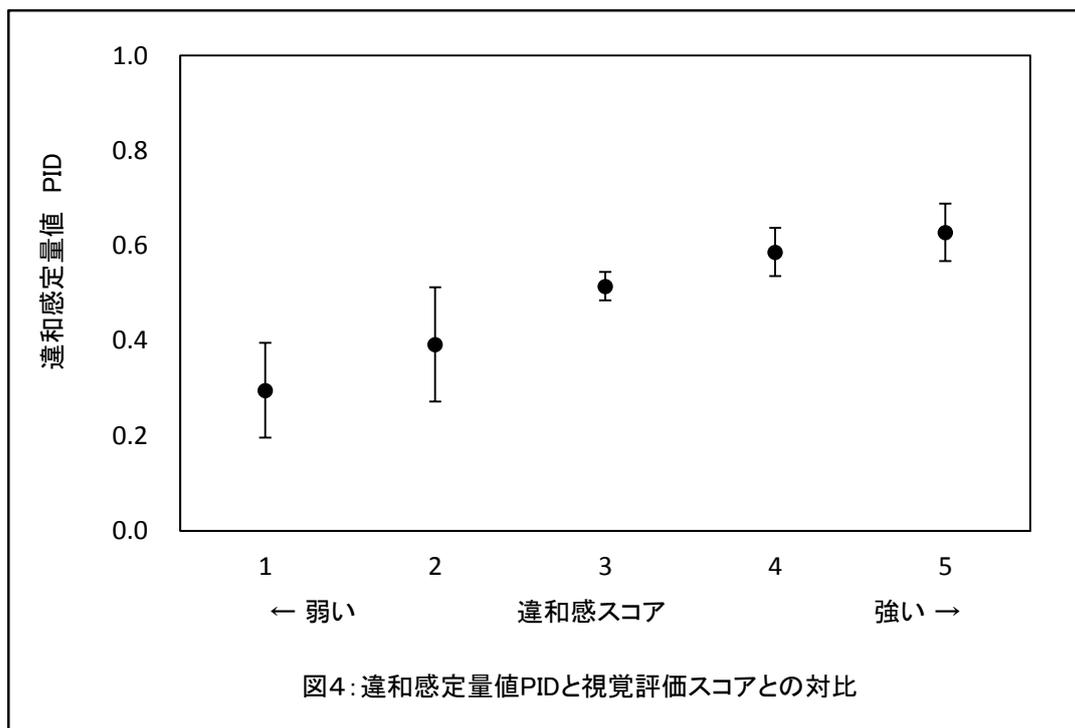
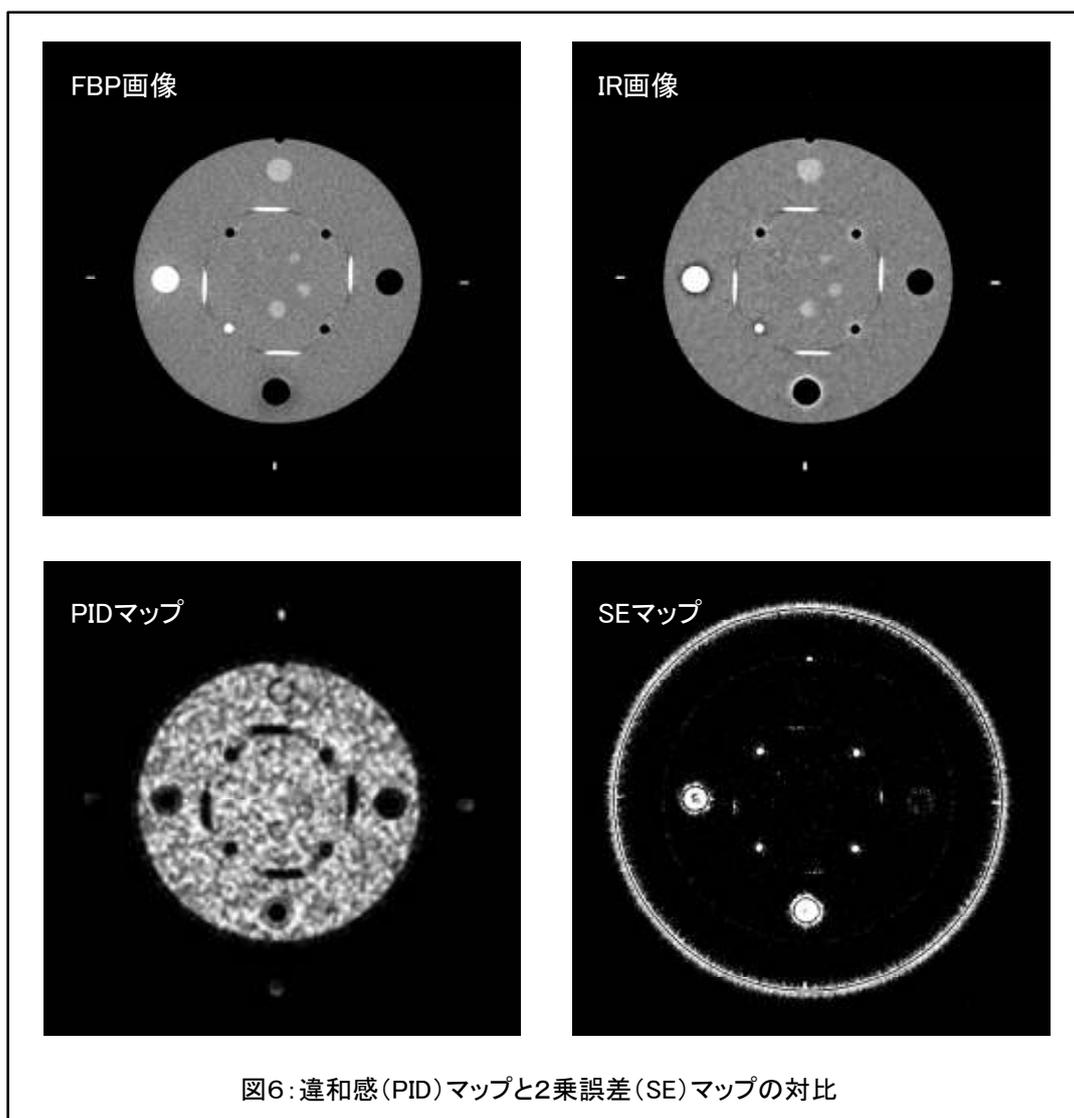


図4より、本研究で提案した違和感 PID は視覚評価にもとづく違和感の強さと矛盾しない定量値を与えることが確認できた。また、評価はノイズレベルの等しいFBP画像・IR画像対で行っており、ノイズレベルの差がPIDや視覚評価結果に影響を与えている可能性は完全に否定できる。したがって、PIDは視覚的印象を反映した新しい画質評価指標となりうることを示唆された。一方、最近注目を集めているMSEは視覚評価とは一致しないことも示された。この理由は、MSE算出過程で得られるSEマップを構造物と対比することで明らかとなった。図6に多目的

モジュールのFBP画像, IR画像, PID算出の過程で得られるPIDマップ, およびMSE算出の過程で得られるSEマップを対比する. PIDマップ, SEマップとも, 高輝度なほどFBP画像とIR画像の相違が強いと判定する. 当該ファントムモジュールには, 水濃度に近いベースに低コントラスト構造物と非常に高コントラストな構造物が存在するが, PIDマップではベースおよび低コントラスト構造物で違和感が強いと判定されているのに対し, SEマップでは高コントラスト構造物やその境界にFBP画像・IR画像間の相違があると判定されている. これは, 高コントラスト構造物ではCT値がベースに対して非常に高く, または低く, 当該領域ではFBP画像, IR間のCT値の差もベースに比べると大きい. その差はさらに平方されることでより強調され, それが画像全体での平均値MSEを支配することになる. 一方, 視覚的にはIR画像の違和感はベース全体に認められ, さらに貴学会理事会からのご意見にもあったとおり, 低コントラスト領域にも強く現れる. したがって, 視覚的印象を反映した画質評価指標はベースあるいは低コントラスト領域でのテクスチャの変化に重きを置かれるべきであり, MSEは不利となる.

本研究は, 主観的な違和感を定量するための新しい指標PIDを提案し, 視覚的評価と矛盾しないことを示した. 同様の研究は国内外で報告されておらず, 本研究はIR法の画質評価において, きわめて斬新かつ重要な知見を与えるものと考えられる.



## 6. 発表等

- (1) 日本医学物理学会機関誌「医学物理」へ投稿中1件.
- (2) 第107回日本医学物理学会学術大会へ演題を提出し受理され, 発表予定.

演題名: 逐次近似法を応用した CT 再構成画像における画質の「違和感」マッピング

## 7. 今後の課題と展開

本研究は, X 線 CT のみならず, 他のモダリティにも応用可能である. たとえば核医学領域においても, IR 法が広く用いられるようになっており, 従来の FBP 画像との視覚的印象の違いがときに問題となる. そこでわれわれは, 核医学断層画像に対しても PID の評価を検討している. さらにその他のモダリティにおいても, さまざまな画像再構成法, 画像処理法が開発されており, 古典的な方法では評価困難な画像の質感の違いなどを評価するのに本研究が役立つと考えている. ただし, PID にはいくつかのパラメータが存在し, これをモダリティに応じて最適化することが必要で, おそらくはX線CTにおいても, さらに最適なパラメータが存在するものと考えられる. これについては現在も研究を進めている.

視覚的評価において問題となったのは, 違和感というきわめて主観的な印象を5段階のスケールで表現することの難しさである. そこで, 同じく主観的で評価の難しい感覚「痛み」の数量化に用いられる Visual Analogue Scale (VAS) の応用を考えている. これは, 疼痛を有する患者が自己の痛みを連続スケールで表現し, 医療者側に提示するためのツールとして広く用いられているもので, ポータブル端末に実装することで容易に視覚評価にも応用可能で, 観察者への負担も少ない. われわれは VAS の応用も含め, よりよい視覚的評価の方法も検討を行っているところである.

## 8. 研究費の使途

<略>

## 9. 特記事項

研究代表者は今年度より, 大阪大学に博士後期課程の社会人大学院生として入学しており, 本研究につきましても, 医用画像情報解析学をご専門とされる村瀬研也先生の研究指導を受ける機会を得ております. 本来であれば, その旨を前年度研究課題報告書に記載すべきでありました. しかし, そのことを失念いたしましたがために, 貴学会理事会より, 大阪大学の松本政雄先生にご指導いただくことをご提案いただきました. 研究代表者の不手際にて大変ご心配をおかけいたしましたこと, 深くお詫び申し上げますとともに, 大変ご多忙中にもかかわらず, このようにご高配を賜りましたこと, 厚くお礼申し上げます. 最後に, 本研究を2年間にわたって助成いただいたことにつきまして, 日本医学物理学会会長 遠藤真広先生, 副会長 松本政雄先生をはじめ, 関係各位に深くお礼申し上げます.