

ケモメトリクスを用いたテラヘルツ分光イメージデータ解析と肝癌組織への適用

保科宏道

独立行政法人 理化学研究所 テラヘルツイメージング研究チーム

〒980-0845 宮城県仙台市青葉区荒巻字青葉 519-1399

Chemometrics in Terahertz Spectral Imaging and Application to Cancer Tissue Analysis

Hiromichi Hoshina

Terahertz Sensing and Imaging Laboratory, RIKEN

519-1399 Aramaki-Aoba, Aoba-ku, Sendai, Miyagi, 980-0845,

Japan

Synopsis

Terahertz spectroscopic images of paraffin embedded cancer tissues have been measured by a terahertz time domain spectrometer. For the systematic identification of cancer tumors, the principal component analysis and the clustering analysis were applied. In three of four samples, the cancer tissue was recognized as an aggregate of the data points in the principal component plots. By the agglomerative hierarchical clustering, the data points were well categorized into cancer and the other tissues. This method can be also applied to various kinds of automatic discrimination of plural components by terahertz spectroscopic imaging.

Keywords: Terahertz time domain spectroscopy, Terahertz spectroscopic imaging, Chemometrics, Cancer diagnostics

1. はじめに

テラヘルツ (THz) 波とは、赤外光とミリ波の中間領域の周波数約 $0.1\text{THz} \sim 10\text{THz}$ (波長 $3\text{mm} \sim 30\mu\text{m}$) の電磁波である。THz 波は電波的な物質透過性を有する最短波長領域であるため、プラスチック・ビニル・紙・繊維・木材・セラミック・半導体・歯・骨・脂肪・乾燥食品・ゴム・コンクリートなどを透過可能であり、X 線に比べて安全なイメージングツールとして使用できる^{1)・4)}。また電波に比べて直進性が高く、光のようにレンズやミラーによる引き回しが可能であり、その透過イメージでは波長程度の適度な空間分解能が得られる。一方でテラヘルツ領域の吸収スペクトルには分子の振動や回転によるモードや、分子間振動、フォノンなどに起因する様々な吸収が観測される。特に粉体の糖やアミノ酸など、結晶性有機物質の多くはテラヘルツ帯に分子間相互作用起源と見られる特徴的な吸収スペクトル(指紋スペクトル)を示すことが知られており、そのスペクトル構造を利用した物質同定が可能である。このため THz 分光イメージからは測定対象の形状だけでなくその物性に関する情報も得ることができる。

近年フェムト秒レーザーの簡便化と低価格化に伴い、それらを用いた THz 波の発生および検出技術が急速に発展してきた³⁾。特にフェムト秒のパルス光をプローブとして用いることで熱揺らぎの影響が格段に小さくなり、従来使用されてきた極低温検出器が不要になったことが THz

波の応用可能性を広げたといえる．現在では空港でのセキュリティチェックや郵便物の未開封検査，生産ラインでの品質検査など様々な分野でテラヘルツ光を用いたアプリケーションが実現しつつある^{1),4)}．医学関連分野では癌と正常組織の吸収差が報告されており，白人で発症率の高い皮膚癌や病理サンプルに対して，複数波長による差分イメージングや分光イメージングによる癌部位の識別が試みられている^{5)~7)}．THz 分光イメージング技術の病理診断応用には，迅速な測定・診断による早期発見，光学顕微鏡に比べて広い領域を測定することによる見落としの低減，医師のキャリアやスキルに依存しにくい客観的識別の実現などが期待される．

一方で THz 波のスペクトルには様々な物質の情報が重なっているために，スペクトルの構造が複雑でわかりにくくなってしまうという問題点がある．実際に THz 帯の分光スペクトルを測定してみると物質の固有モードによる吸収だけでなく，結晶多形やサンプル形状による Mie 散乱，パッケージの干渉など様々な効果の重ね合わせとして得られてしまい，その解析が単純でない事が多い^{8),9)}．特に生体組織の THz 波吸収スペクトルは特徴的な吸収ピークを示さないため，組織構造の複雑性，個体差，病変部位による差異や癌の種類等の様々な条件への対応が必要であり，癌組織を選択的に抽出することは容易でなく，従来行われてきた吸収率やスペクトルの単純な比較に代わる新しい分

析手法が必要である．

そこで本研究では，THz 分光イメージを用いた異質部位の系統的識別法の確立を目的として，癌組織の THz イメージの解析に対するケモメトリクス¹⁾の適用を試みた．具体的にはパラフィン包埋した肝癌組織切片の THz 時間領域分光イメージに対して主成分分析とクラスタ分析と組み合わせ，癌部位の分離を試みた．その結果，4 つのサンプルのうち 3 つの試料で癌部位が分類され，光学顕微鏡観察とのよい一致が確認された．このような異質部位の弁別は様々な応用やシステム開発でも重要であり，本研究はそのひな形を提供すると期待される．

2. THz 分光イメージングシステム

本研究における THz 分光イメージの測定には，THz 時間領域分光（THz-TDS）を利用した THz 透過分光イメージングシステム（（株）栃木ニコン）を用いた．THz-TDS は，フェムト秒レーザーによって発生させた単パルスの THz 波の電界をプローブ光で直接検出し，その時間波形のフーリエ変換から THz 波の振幅と位相スペクトルを得る手法である^{1)・3)}．サンプルとリファレンスの振幅，位相スペクトルからサンプルの吸収係数および屈折率のスペクトルが得られる³⁾．

システムの概要を図 1 に示す．フェムト秒レーザー光はビームスプリッターで励起光と検出光に分離される．

図 1

励起光は，低温成長ガリウムヒ素（LT-GaAs）基板上のギャップ幅 $100\mu\text{m}$ の 2 つの電極間に照射される．電極間には数 10V のバイアス電圧が印加されている．フェムト秒レーザーパルスで基板上に発生した光キャリアが電極間電場によって加速されることで瞬時電流が流れ，その時間微分に比例した電界振幅を有する THz 波が発生する．発生した THz 波は基板に付けられたシリコン超半球レンズと放物面鏡を用いていったんサンプルチャンバー内で集光され，その後に対称な光学系を通して検出素子上に集光される．図 1 のシステムでは，ギャップ幅 $5\mu\text{m}$ の 2 つの電極が作製された LT-GaAs 基板を検出素子として用いている．検出用のフェムト秒レーザー光は，光学遅延ステージを経由して検出素子に集光される．フェムト秒レーザー光が 2 つの電極間に照射されると，放射素子と同様に光キャリアが生成されるが，電極間には電位がかかっていないため，通常は電流が流れない．しかし，キャリア発生時に THz 波が入射すると，その電界により瞬時電流が流れるため，その電流値を検出することで光パルスが照射された瞬間の THz 波電場を測定できる．通常，この光パルスは THz 波電場の時間変化より十分に短いため，遅延ステージを用いて検出用レーザーパルスと THz 波パルスの入射時間差を変化させることにより，THz 波電場の時間波形が得られる（図 2）．実際のシステムでは検出感度を高めるためにバイアス電圧に変調をかけ，ロックイン検出をしている．

THz-TDS 法による分光イメージング測定のためには、サンプルチャンバー内の THz 波の集光位置にサンプルを設置し、X-Y 軸ステージによってサンプルを移動させることで 1 画素ずつ THz 波の時間波形を取得する。それぞれの位置での透過時間波形に対して高速フーリエ変換 (FFT) を施すと、その位置のサンプルを通過した THz 波のパワースペクトルと位相スペクトルを取得できる。得られたスペクトルはサンプル無しで測定したリファレンススペクトルを用いて³⁾吸光度スペクトルと屈折率スペクトルに変換される。ここで吸光度および屈折率の導出に必要なサンプル厚は、サンプル上で測定した数点の厚みの平均値を用いている。なお、空気中の水蒸気による吸収の影響を低減するため、本測定ではいったんサンプルチャンバー内を真空引きしてから窒素ガスを導入して測定を行っている。

3. THz 分光イメージングの結果

本研究では、測定サンプルとしてパラフィン包埋処理された 4 つの無染色肝臓組織切片（厚さ約 1mm）を用いた。THz 透過分光イメージは、周波数分解能 25GHz、積算回数 40 回、画素サイズ 250 μ m で計測した。計測時間は 1 サンプルあたり約 7 時間である。解析には装置が十分な S/N を持つ周波数範囲である 0.5THz $\sim$ 2.25THz のスペクトルを用いた。図 3 左端に 4 つの肝癌組織切片の光学写真を示す。赤い斜線部は、同じ組織ブロックから数 10 μ m の

図 3
カラー

厚みで切り出した別サンプルを染色して光学顕微鏡観察により医学的に同定された癌組織領域を示す．図 3 の中央及び右に，THz イメージの例として 1.7THz における吸光度及び屈折率画像を示す．図 3 の THz イメージを見るとサンプル A 及び C では吸光度画像で，サンプル B では屈折率画像で癌領域と他の組織領域が識別できていることが確認できる．このような傾向は図 3 に例示した 1.7THz の周波数だけでなく，0.5THz～2.25THz の全ての周波数領域で確認できた．

これらのスペクトル上の癌領域の特徴は，サンプルにより大きく異なっている．図 4 に 4 つの肝癌組織の癌領域及び癌以外の領域から抜き出してきた吸光度及び屈折率のスペクトル例を示す．全てのサンプルのスペクトルで吸光度，屈折率ともに特徴的なピークが見られないことがわかる．従ってスペクトル上で組織の違いを判別するためにスペクトル強度や傾きなどに頼ることになる．しかし，サンプル A の吸光度スペクトルでは癌の吸収が癌以外の吸収より大きい，サンプル B，C，D では逆になっている．また，屈折率スペクトルでは癌組織はサンプル B でのみ確認され，屈折率の値自体もサンプルや測定位置によって大きく変動している．

図 4

このような生体組織の THz スペクトルの違いが何に起因する物か，現在のところ明らかになっていない．水の THz 帯における吸収の強さを考えると，おそらくサンプ

ルの吸光度の変化は組織の水分含有量によるものであると予想できるが，その他にも組織形状による散乱の効果など無視できないものも多く，得られたスペクトルの物理的起源を追求するのは容易ではない．しかし得られた THz 分光イメージからは癌部位とそうではない部位の間にコントラストが認められることから，それらを系統的に識別する方法を開発することで，実際の癌診断に役立てられる可能性がある．そこで我々は多変量で複雑な系の検量や分類に使うことができ，近赤外領域のスペクトル解析などで実績をあげているケモメトリックスを THz 分光イメージデータの解析に導入し，自動的・系統的な癌の識別を試みた．

4. ケモメトリックスの適用

4.1. 主成分分析

ケモメトリックス (Chemometrics) は，Chemistry (化学) と metrics (計量法) を組み合わせた造語であり，スペクトルのような化学分野の多変量データに対して数学的・統計学的手法を活用して，最大限の情報を引き出す解析のことを表す¹⁰⁾．本研究では主成分分析¹¹⁾とクラスタ分析¹²⁾を組み合わせる THz 分光画像の解析処理に用いた．なお実際のデータ処理には MINITAB14 (Minitab, Inc.) を使用している．

本研究では，まずケモメトリックスの第 1 段階として主成分分析を行った．主成分分析とは，スペクトル内のサ

サンプルの特徴や差異の情報を，より効率的に明確に検出することを目的とした解析方法である．図 5(a)，(b)，(c)に主成分分析の概要図を示す． n 個の周波数データ点で構成される吸光度または屈折率スペクトル $f(\nu)$ がある場合（図 5(a)），イメージ上のある 1 点のスペクトル情報 A は，周波数 ν_i ($i=1,2,\dots,n$)でのスペクトル強度 ($f(\nu_i)$)を軸とする n 次元空間内の 1 点として表現できる(図 5(b))．同一サンプル内の全ての画素の THz スペクトルをこの n 次元空間にプロットすると，各画素のスペクトルの違いが空間上の距離として現わされる．最も顕著に特徴が表れる現れる方向，すなわち最もデータ点の分散が大きい方向に第 1 主成分軸，2 番目に分散が大きい方向に第 2 主成分軸，と新たな軸を再定義することによって多次元のスペクトル情報を低次元の主成分空間上に射影することができ，各データ点の類似度や相関関係を効率よく抽出できる（図 5(c)）．このような解析をする上でどの主成分軸(固有ベクトル)までを解析に用いるかは，主成分軸のローディング（固有値）を参考にして決定するが，本研究では第 1 および第 2 主成分によりほとんどのデータが説明できる事が判明したため，その 2 成分を解析に用いている．

図 6 は測定したサンプル A~D のスペクトル主成分分析にかけ，各画素のスペクトルを第 1 主成分軸と第 2 主成分軸で作られる平面内にプロットしたものである．プロットの中の 1 点がサンプルの 1 画素のスペクトルを表してお

図 6
カラー

り，グラフ内の赤点が癌組織，黒点がそれ以外の領域の画素に対応している．１列目は吸光度スペクトルを用いた主成分プロットであり，２列目は屈折率スペクトルを用いたものである．３列目は吸光度と屈折率の両方のスペクトルデータを用いた主成分プロットであるが，主成分分析を行う上で吸光度と屈折率の重みが主成分空間内で同等となるようにあらかじめデータの標準化がされている．

吸光度及び屈折率イメージの片方を用いた主成分プロットを見ると，癌組織の分離の可否はサンプルによって異なることが見て取れる．これは図３の吸光度画像，屈折率画像で見たようなサンプルごとの癌組織のスペクトルのばらつきが主成分プロットにもそのまま現れていることを意味している．これに対し，吸光度と屈折率の両方の THz 分光イメージングデータを用いた場合には，サンプル A，B，C で黒点の癌組織と灰色点の癌以外の組織がより明確に分離されていることがわかる．両方のスペクトルを同時に処理することで，その中に含まれる情報を効率的に引き出すことができたと思われる．さらに細かく見ると，組織が似通っているほど点と点が近くに分布していることも確認できた．そこで，点と点との距離に着目して自動的なグループ分けを実行する手法としてクラスタ分析を適用し，癌部と癌以外の部位との弁別を試みた．

4.2. クラスタ分析

クラスタ分析とは距離を基準に空間内の多数のデータ点を複数のグループに分類する手法である。本研究では、このクラスタ分析を主成分プロット内のデータ点の類別に利用した。図 5(d)にクラスタ分析の概念図を表す。図 5(d)は、図 5(c)の 2 次元主成分空間内での距離を基準に最終的に 2 つのグループに分類した場合に対応している。

クラスタ分析には、階層的手法と非階層的手法がある。非階層的手法の代表的な分類法として **k-means** 法が挙げられる。**k-means** 法は、ある決まった数のクラスタに分類する方法である。例えば、サンプルを **k** 個の集団に分類する場合には、通常は **k** 個の初期集団の平均値をそれぞれ初期値に与えることが多いため **k-means** 法と呼ばれる。この **k-means** 法では、最初に代表となるデータ点を選ぶ必要があり、その種から他の点までの距離が最小となる様にデータ点の出し入れを繰り返しながらグループ分けを収束させて分類する。その結果、最初に選ぶ点の取り方によってクラスタ分析の結果が左右される。このように **k-means** 法は、あらかじめ代表的な開始点が既知であるか、もしくは、指定したいグループが存在する場合に有効である。

一方階層的手法とは、最初に最近接の 2 点からなる多数の小クラスタを形成させ、次にそれらの小クラスタ間の距離を基準にして、近傍の小クラスタを次々に統合していく方法であり、最終的には全ての点が 1 つのクラスタと

して形成される．このため，いったん形成されたクラスタから点が離脱したり入れ替えられたりすることはなく，多数の分木の組み合わせによって階層的にクラスタへの帰属関係が表現できる．クラスタ統合の計算方法としては，「平均法」，「最短距離法」，「メジアン法」などがある．本研究では様々な距離空間の計算方法を主成分プロットに適用させた結果，より癌が明瞭に分類された階層クラスタ分析（Ward 法）¹²⁾を採用した．

図 7 は 4 つの肝癌組織切片に非階層クラスタ法（k-means 法）と階層的クラスタ法（Ward 法）を適用した結果を示す．それぞれの結果を主成分プロット（左側）と実空間でのイメージ（右側）で示した．k-means 法を適用するにあたって，開始点として病理診断であらかじめ得られた癌の領域と癌以外の領域からそれぞれ一点選んでいる（図中の青い×印）．一方で階層的クラスタ法を適用する際には，病理診断であらかじめ得られた癌部位の情報は必要でないため，予断を挟むことなくクラスタ分けすることができた．ただし，得られた領域のどちらが癌に相当するか判断することは本質的にできない．

k-means 法の結果を見ると，主成分プロット上で直線上に領域わけされていることがわかる．そのため，実際の癌組織の分布を反映せず，実空間イメージでも癌組織の領域と異なったクラスタ分けになっている．非階層クラスタ分析を用いると，このような単純なクラスタ構造になっ

図 7

2 段使って大きく表

示（横 16 c m）

カラー

てしまう事が多く，病理サンプルで同定された癌組織の領域を反映した結果が得られなかった．

一方，階層的クラスタ分析である Ward 法を適用した結果を見ると，サンプル A, B, C で主成分プロット上のクラスタ分けが図 6 の癌部位と癌以外の部位の分類と似通ったものになっている．また実空間イメージ上でも光学写真内の斜線で示す癌領域とクラスタ分けの結果がよく一致している．なかでもサンプル A と B では一致率が高く，癌部位と癌以外の部位の識別に成功していることがわかる．一方，サンプル C ではサンプル中央にひび割れがあったため，最終的に癌組織・癌組織以外・サンプルの欠陥部分の 3 つのグループに分類した．また，図 3 の THz 分光画像からも識別が不十分だったサンプル D は，ケモメトリクスを用いても識別はうまくできなかった．これは他のサンプルよりもサンプル D の癌領域内で癌組織が壊死している部分が多いことが要因のひとつであると想像されるが，医学的な考察も含めたより詳細な検討が必要であろう．

5. まとめ

本研究では，4 つのパラフィン包埋された肝癌組織を用いて透過型 THz 分光イメージングを行ない，それぞれの吸光度イメージデータ及び屈折率イメージデータを測定した．得られた THz 分光イメージデータから癌組織の情報を抜き出すためケモメトリクスを適用した．最初に吸光度

と屈折率の分光イメージデータに対して同時に主成分分析を行った結果，3つのサンプルで癌と癌以外の部位が主成分空間上で明瞭に分離しているのが確認された．次に主成分プロットにクラスタ分析を適用し，異質な成分の自動的な分類を行った．その結果，3つのサンプルにおいて，得られた分類イメージと病理診断結果が一致する結果が得られ，本手法の有効性が確認された．

本手法は，あくまでサンプル内の異質な領域を弁別するものであるため，実際の病理診断に適用する際には医師による光学顕微鏡下の観察・検証が必要である．しかしTHz光は可視光での検査や，ラマン分光⁽¹³⁾や赤外吸収分光^(14,15)を使った癌診断に比べてサンプルに対する透過性が高いため，厚さ1mm以上の厚いサンプルの測定が可能である．また光学顕微鏡に比べて視野も広いことから広範囲の測定が可能で，医師の診断の手助け（客観的な診断）や見落とし防止に利用可能な解析方法として有効である．今後は迅速な診断を可能にするため，THz光が氷を透過することを利用して，冷凍組織の癌診断を試みる予定である．

最後にTHzイメージへのケモメトリクス⁽¹⁶⁾の適用は癌組織の同定をはじめとする医学応用に限定されるものではなく，他のTHz分光画像に対しても有効であり，様々な応用分野において複雑な多変量THzスペクトルイメージデータを系統的に分類する新しい解析方法として期待できる．

謝 辞

本研究は理化学研究所テラヘルツイメージング研究チームと福井大学医学部の共同研究で行われたものである。肝癌サンプルを提供していただいた三好憲雄博士，サンプル測定及び解析に多大な貢献をしていただいた中島佐知子さん，実験とディスカッションにご協力いただいた山下将嗣博士，そして大谷知行チームリーダーに深く感謝いたします。また，本研究を立ち上げるきっかけをつくっていただいた名古屋大学の川瀬晃道教授に深く感謝いたします。

図の説明

図 1

THz 分光イメージング装置の模式図.

図 2

THz-TDS によって測定した THz 波の時間波形及びそのフーリエ変換後のスペクトル.

図 3

肝癌組織の光学写真, および 1.7THz における吸光度画像と屈折率画像.

図 4

肝癌組織の吸光度及び屈折率スペクトル例.

図 5

主成分分析及びクラスタ分析の概要. (a) スペクトルデータ, (b) 主成分分析, (c) (a)の主成分分析結果(主成分プロット), (d) (a)のクラスタ分析結果.

図 6

肝癌組織切片の THz 分光スペクトルから作成した主成分プロット上での癌と癌以外の組織の分布.

図 7

肝癌組織切片の光学写真と非階層クラスタ法(k-means法)と階層的クラスタ法(Ward法)を適用した結果. それぞれのクラスタ法で左側の図が主成分プロット, 右側の図が実空間イメージ上での結果を表している.

参考文献

- 1) D. L. Woolard, W. R. Loerop and M. S. Shur: *Terahertz Sensing Technology: Emerging Scientific Applications & Novel Device Concepts* (World Scientific, Singapore, 2004).
- 2) 酒井清美, 廣本宜久: 分光研究 第 54 卷 第 1 号 43 (2005)
- 3) 萩行正憲: 分光研究 第 54 卷 第 3 号 181 (2005)
- 4) 川瀬晃道, 小川雄一, 大谷知行: 分光研究 第 54 卷 第 5 号 314 (2005).
- 5) R M. Woodward, B. E. Cole, V. P. Wallace, R. J. Pye, D. D. Arnone, E. H. Linfield and M. Pepper: *Phys. Med. Biol.* **47** 3853 (2002).
- 6) P. Knobloch, C. Schildlkencht, T. Kleine-Ostmann, M. Koch, S. Hoffmann, M. Hofmann, E. Rehberg, M. Sperling, K Donhuijsen, G. Hein and K. Pierz: *Phys. Med. Biol.* **47**, 3875 (2002).
- 7) P. Tady, I. Bradley, D. Arnone and M. Pepper, *J. Pharm. Sci.* **92**, 831 (2003).
- 8) C. F. Zhang, E. Tarhan, A. K. Ramdas, A. M. Weiner and S. M. Durbin, *J. Phys. Chem. B*, **108**, 10077 (2004).

- 9) E. Pickwell, B. E. Cole, A. J. Fitzgerald, M. Pepper
and V. P. Wallace: *Phys. Med. Biol.* **47**, 1595 (2002).
- 10) 尾崎幸洋, 宇田明史, 赤井俊雄: 化学者のための多
変量解析 (講談社, 東京, 2002) .
- 11) J. N. Miller, and J. C. Miller: *Statistics and
Chemometrics for Analytical Chemistry* (Prentice
Hall, New York, 2001).
- 12) J. Ward: *J. Am. Stat. Assoc.* **58**, 236, (1991).
- 13) D. C. B. Redd, Z. C. Feng, K. T. Yue, and T. S.
Gansler, *Appl. Spectrosc.* **47**, 787 (1993).
- 14) Y. X. Ci, T. Y. Gao, J. Feng, and Z. Q. Guo, *Appl.*
Spectrosc. **53**, 312 (1999).
- 15) P. Lasch, W. Haensch, D. Naumann, M. Diem,
Biochem. Biophys. Acta, 1688, 176 (2004).

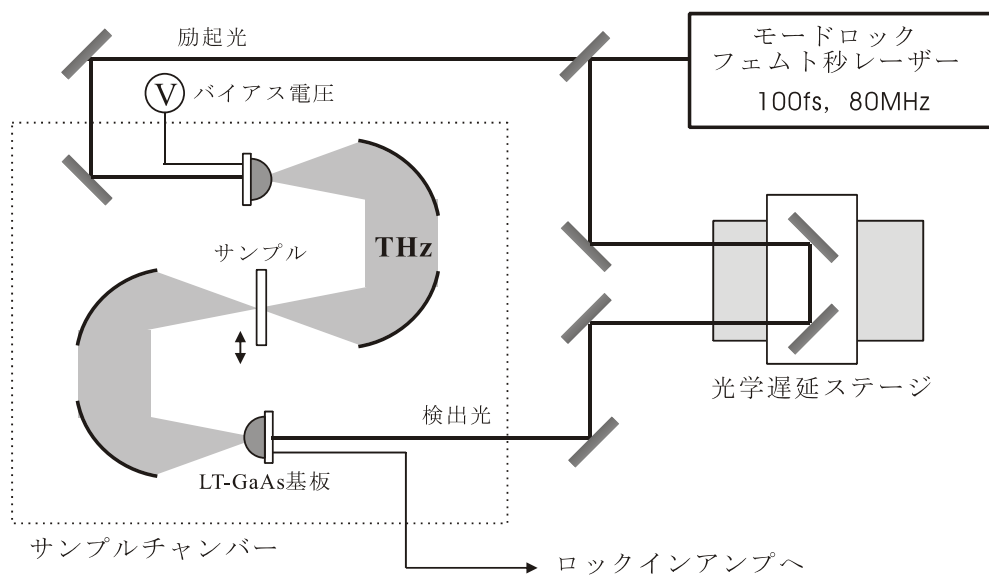
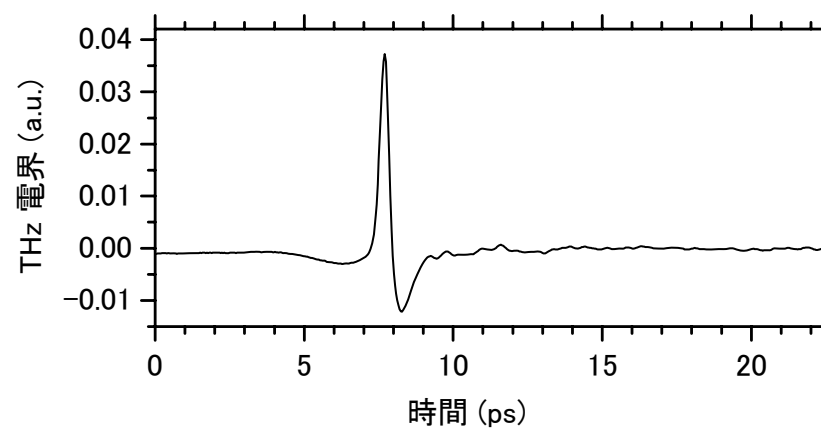


図 1

保科宏道



↓ フーリエ変換

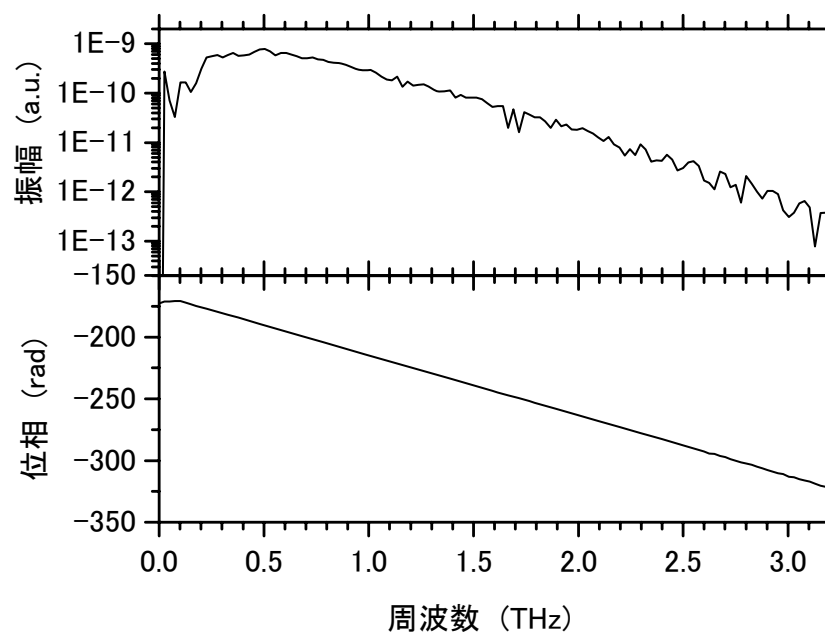


図 2

保 科 宏 道

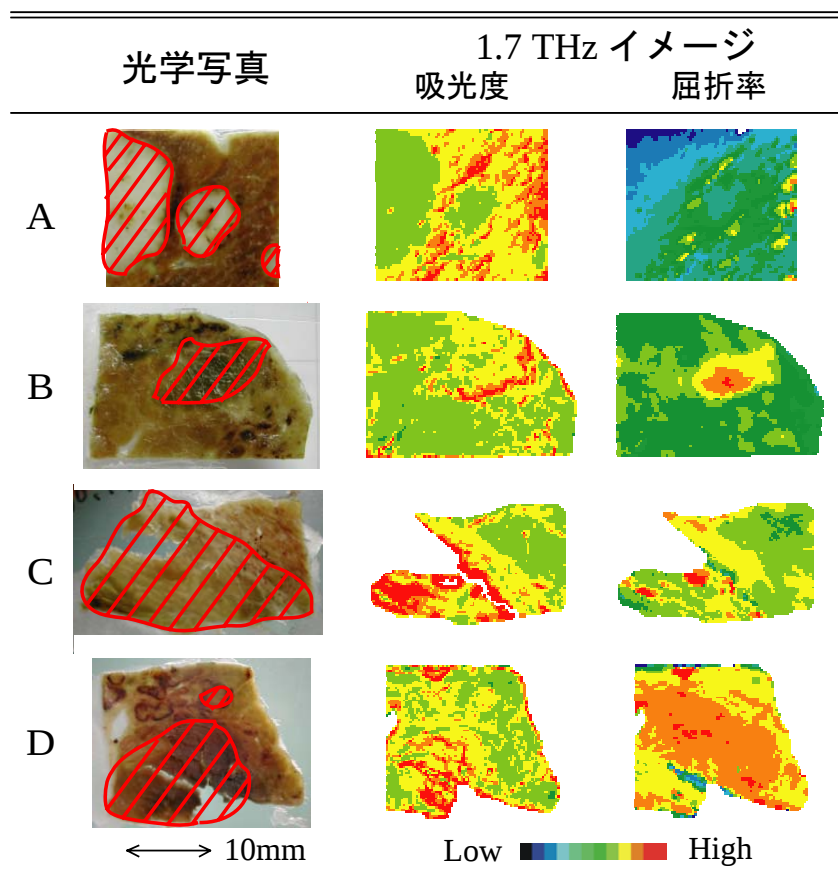


図 3 ,カラー印刷

保科宏道

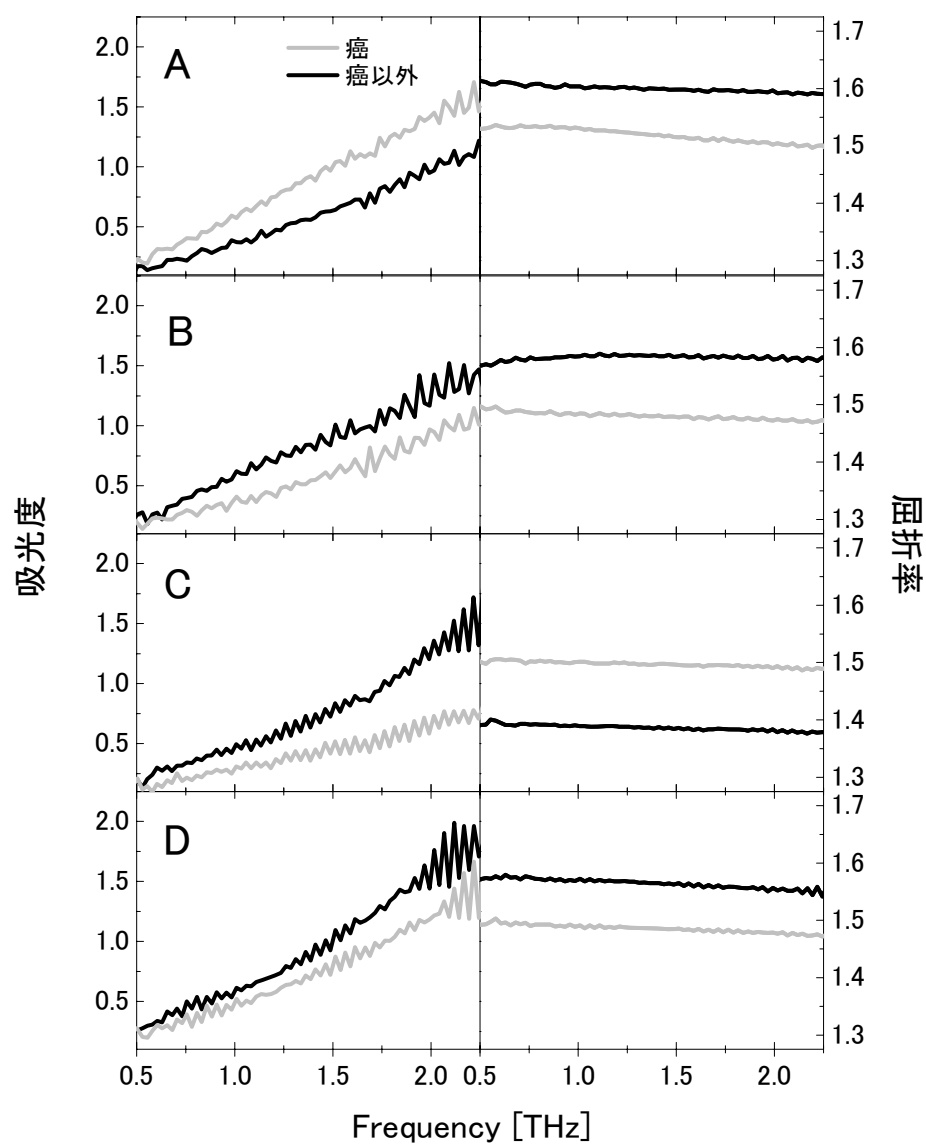
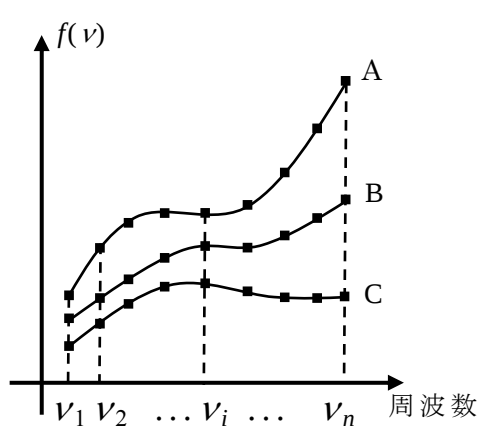
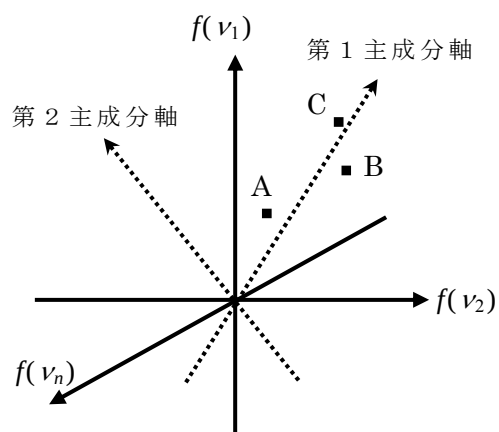


图 4

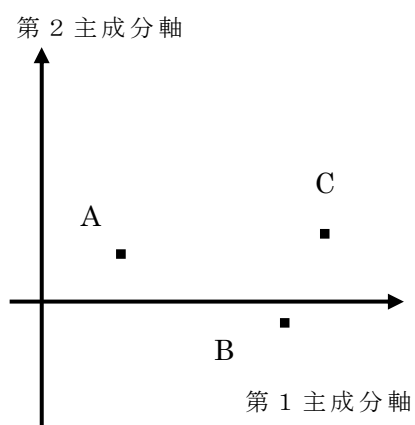
保科宏道



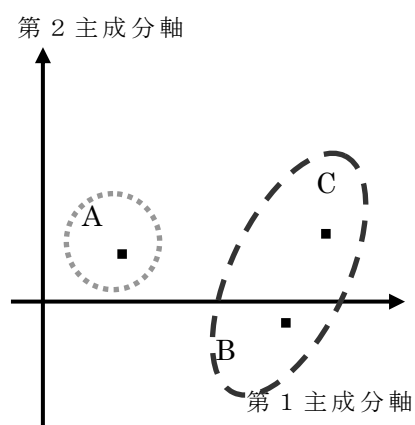
(a)



(b)



(c)



(d)

図 5

保科宏道

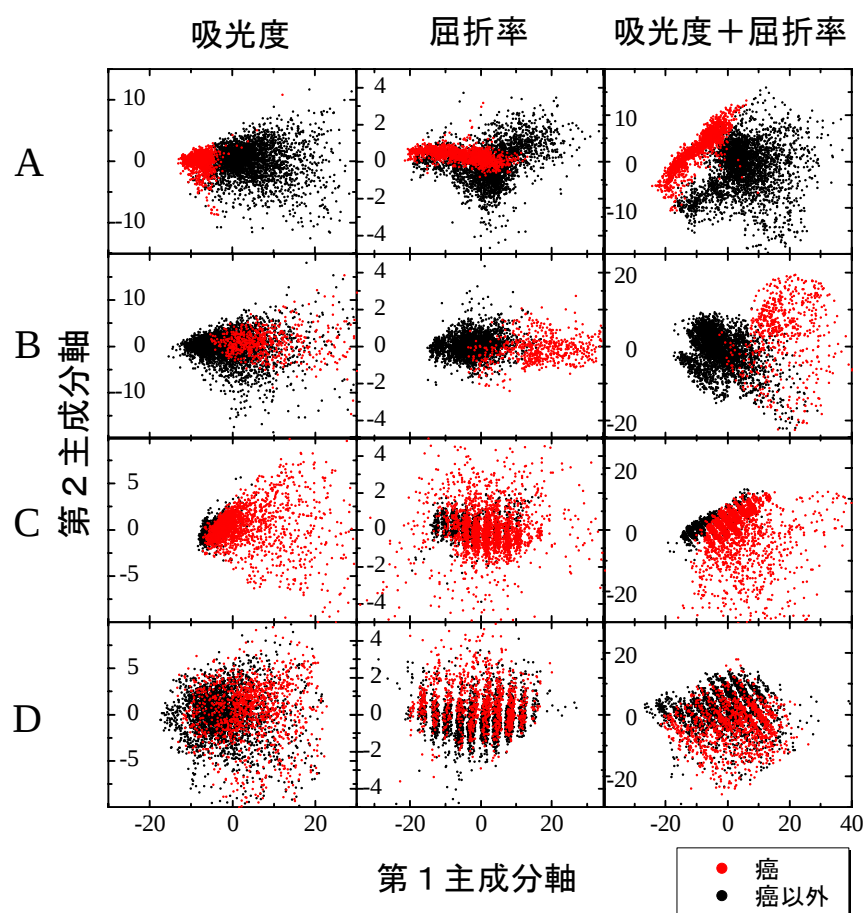


図 6 , カラー印刷

保科宏道

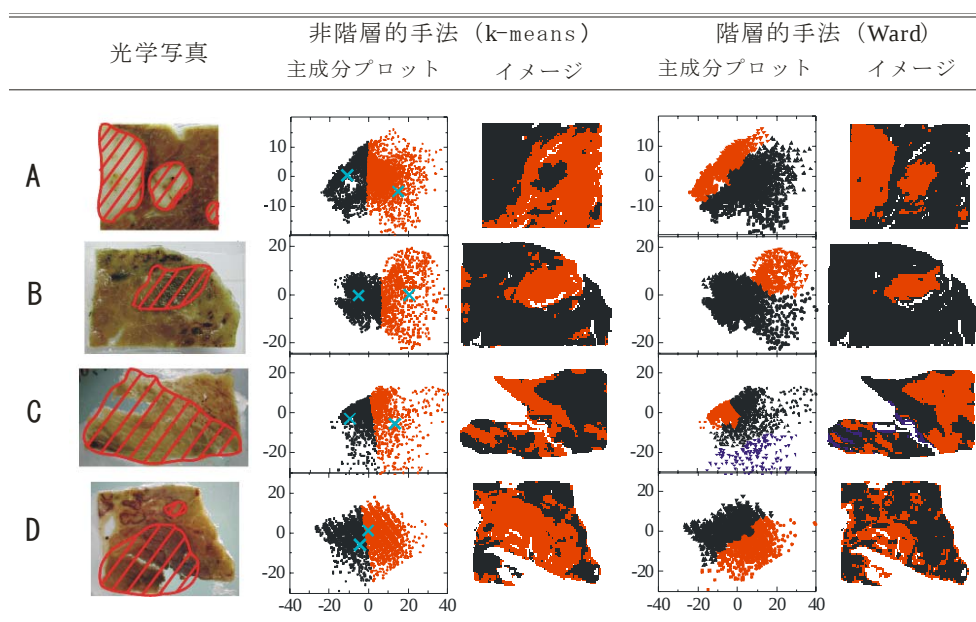


図 7 , カラー印刷

保科宏道