

医用画像における特徴の可聴化を利用した 画像診断支援手法の開発

賀川 経夫^{1,a)} 田上 秀一^{†1} 清末 一路^{†1} 森 宣^{†1} 西野 浩明¹

概要: 近年, 医用画像を取り巻く技術が急速に進歩し, CT や MRI など放射線等を利用して身体の内部を撮影する手法においては, ミリ以下の単位での断層を高解像度の画像で撮影することが可能である. その反面, 診断の際には, 担当する医師は, 一人の患者あたり膨大な数となるそのような画像を精査しなければならない. 1日に何例も扱う場合は, 数万枚に及ぶこともある. 本研究は, このような負担のかかる画像診断時において, 効果音を利用することにより, 診断者の負担の軽減を目的とする. そこで, 医用画像内の特徴に基づいた効果音を生成し出力するという可聴化処理を行うことにより, 前述の診断者の負担を軽減する手法について検討を行っている. 本論文では, 効果音を利用した支援手法を提案するとともに, テストベッドを用いた評価について述べる.

Development of Medical Diagnosis Support System based on Sonification of Image Features.

TSUNEO KAGAWA^{1,a)} SHUICHI TANOUÉ^{†1} HIRO KIYOSUE^{†1} HIROMU MORI^{†1} HIROAKI NISHINO¹

Abstract: Recently, Computer Aided Diagnosis (CAD) has become one of the most important for medical activity. The more exact and various CAD become, the larger amount of medical images are provided. Furthermore, these images becomes high definition. Radiologists have to cost their time and efforts to investigate these medical images. It is strongly required to reduce their burden without debasing the quality of imaging diagnosis. In this paper, we propose the technique to generate sound information based on the image features and discuss their effects for diagnosis. Generating sound effects helps for attention rousing and the fatigue reduction for medical imaging diagnosis.

1. はじめに

医療分野における画像処理環境の発達とソフトウェアの充実により, CT や MRI などの高画質画像を用いた画像診断が可能となっている. しかしながら, 画像診断においては, 専門医師は異常個所を見つけるために, 大量の高精細画像を全て集中して精査しなければならない. 1日に何例も扱う場合には, 調査する対象画像が何万枚にも及ぶため,

相当の負担を背負うことになり, 疲労による見落としなどの医療過誤につながる恐れがある. このことに対して, 異常個所を検出するために必要な画像特徴を抽出し, 強調提示することにより, 診断の負担を軽減する技術に関する取り組みが多くなされており, 一部の事例に対しては, 大きな成果を挙げている. しかしながら, 病気の種類や進行状況, 疾患の状態によって, 抽出すべき画像特徴が大きく異なるため, 特徴を絞り込むための画像処理の自動化が非常に困難である. したがって, 画像処理技術や処理環境がいくら整備されても, 最終的には診断者の視診によって異常個所に対する診断を下す必要がある.

精細度の高い画像を見続ける医用画像診断システムの構築において, 複数モダリティの利用, すなわち視覚モダリ

¹ 大分大学工学部知能情報システム工学科
Department of Computer Science and Intelligent Systems,
Oita University, Oita, Japan

^{†1} 現在, 大分大学医学部放射線科
Presently with Department of Radiology, Oita University
Faculty of Medicine, Oita, Japan

^{a)} t-kagawa@oita-u.ac.jp

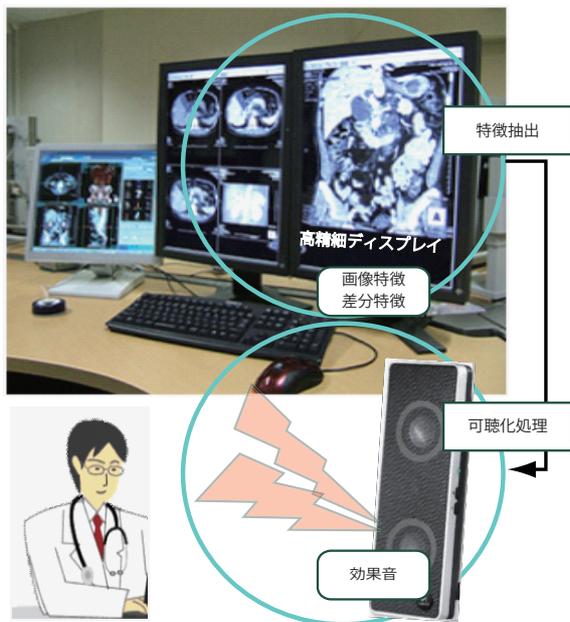


図 1 医用画像診断支援システム

ティである画像に加えて聴覚モダリティである音情報の利用を考える。音の変化によって画像の変化を知らせることができれば、診断者に対する注意喚起となり、「見続ける」ことに関する疲労を低減することができると考えられる。また、効果音の生成に関しては、診断に必要な特徴が音で表現されているため、画像を提示するとき、必要以上の画像強調処理を行わず、できるだけ元の画像を提示することができる。

本研究では、画像診断時に医用画像に関して画像内の特徴に基づいた効果音を生成し出力するという可聴化処理を行うことにより、前述の診断者の負担を軽減する手法について検討を行っている。本論文では、効果音を利用した支援手法を提案するとともに、テストベッドを用いた評価について述べる。

2. 画像の可聴化技術

2.1 画像診断技術

CTやMRIによって得られる連続断面画像による診断では、図1のように高精細ディスプレイ上に診断用の画像を出力し、3次元化や強調処理などの画像処理を駆使して、診断者が患者の疾患やその前兆を発見する。現在、断面画像は、ミリ以下の単位で撮影することが可能であるため、一人の患者の分でも数千枚に及ぶことがある。図2に診断対象となる一連の断面画像の一部を示す。

画像処理技術を用いた画像診断支援手法として、異常箇所における既知の特徴を強調し、その位置を推定する手法が多く提案されている [1], [2]。これらの手法は、画像診断にかかる診断者の時間と負担を減少させるためのものであり、完全に自動的な判断を行うには至っていない。多くの場合、画像強調する手法では、同一手法で安定した処理結

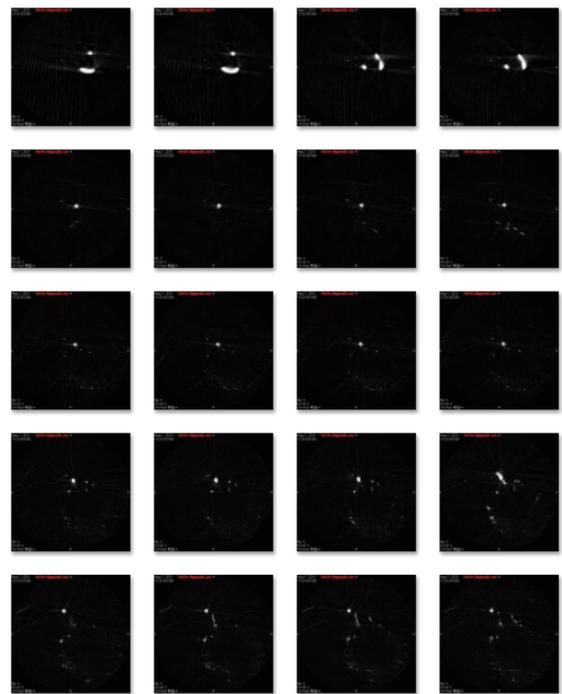


図 2 対象医用画像

果を得るために、正規化等の前処理を的確に行う必要があり、画像処理の精度が大きな問題となる。また、異常でない部分も候補として多く得られることもある。診断者は、異常の有無も含めて最終的に確認を行わなければならない。さらに、診断のために加工された画像が、どのような経緯で何の部分が処理されているのかを診断者が正確に把握する必要がある。したがって、画像処理に関する知識も要求される。

2.2 画像の可聴化技術

従来、視覚障害者を対象とした画像の可聴化を行う研究がすすめられている [8], [9], [10], [11], [12]。可聴化とは、ある情報を効果音や振動に変換し、聴覚モダリティとして提示する技術である。画像の可聴化については、音声合成 (Text to Speech, TTS) を用いて画像内容を自然言語で説明する手法が提案されているが、言語音声化については、用途が限定される、度合いなどの言語化が困難な情報が含まれることがあるなどの問題がある。

また、ビデオ画像に対して、色分布や輝度情報、エッジ情報などの画像の特徴を、一定の規則にしたがって、リアルタイムに音に変換することにより、視覚障害者が音で環境を認識できるように音で状況を知らせる手法などが提案されている。しかしながら、大まかな情報の提示には有用であるが、画像内の詳細な情報の提示には不向きである。また、診断者は、画像内に撮影されたものと出力された音との対応関係をあらかじめ習得し、音声提示に慣れる必要がある。

効果音など他の感覚モダリティを診断の補助として利

用することにより、大量の断面画像を見続けることによる視覚の疲労を軽減し、効率よく診断を行うことができる [3], [4], [5], [6], [7]。また、診断に必要な特徴が音で表現されるため、画像を提示するときに、必要以上の強調処理を行わず、できるだけ元の画像を提示することができる。そのため、診断者の見たい画像を提示できるので、画像処理の詳細な内容や精度などの専門的な知識を特に有する必要がない。

複数モダリティの併用を前提に医用画像に可聴化を行い、診断を支援する手法については、以下の点を検討する必要がある。

- 画像内で顕著に特徴が表れることに関して、相応の効果音を付与しても、あまり効率的ではない。視覚的に識別しにくい画像特徴に対して、より明瞭な効果音を生成し、見落としを起こしにくくする。
- しかし、音を精細に聞き分けなければならないような診断手法にすると、見るだけでなく、聞くことにも集中しなければならず、かえって余計な労力をかけることになる。したがって、効果音にはある程度の冗長性を持たせなければならない。
- 画像特徴に対するイメージと効果音の特徴が一致していない場合、診断者に混乱を招く恐れがある。例えば、注目している領域が大きくなっていくように見えるときに、効果音が高くなったり、小さくなったりすると、かえって違和感が生じてしまい、診断の妨げとなる恐れがある。音と画像とに互いに整合性を持たせ、その関係を明確に診断者に理解してもらえるように提示する必要がある。
- 医用画像の用途や撮影箇所、予想される症状などによって、画像内に表れる特徴が変化し、それに応じて診断者の興味も変化する。したがって、診断者の注意する部分や画像特徴に関して、画像処理に関する知識が乏しくても、その意図を正しく反映し、思い通りに操作できるインタフェースを設計する必要がある。

3. 医用画像診断支援システム

3.1 概要

図3に本手法の処理の概要を示す。現段階では、図4に示すように、血管の断面を主たる対象として、効果音の付与を検討している。図4の左側は、実際の血管を3次元コンピュータグラフィックスで表現したものであり、右側が実際に得られた断面画像である。

このように処理対象であるCTやMRIによる連続的な断面画像のシーケンスが与えられると、256階調の輝度画像に変換される。ついで、以下に示す処理により、一連の断面画像のシーケンスに対して効果音が付与される。

- (1) 画像の特徴抽出
- (2) 効果音の生成

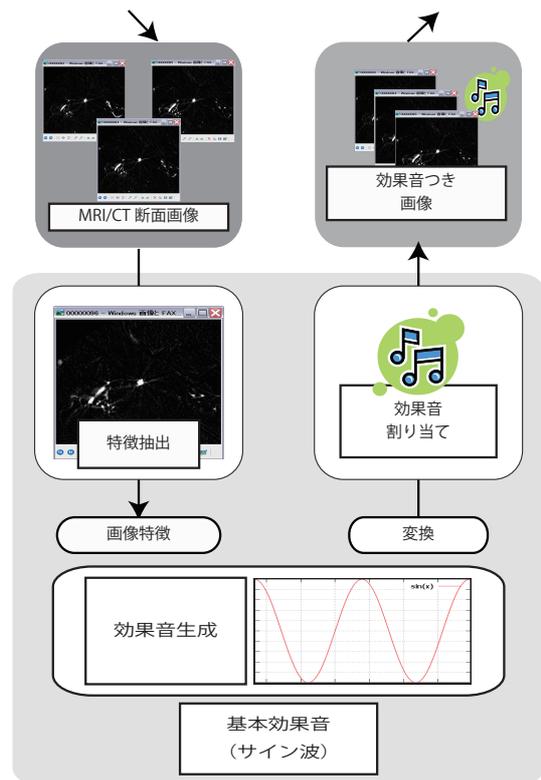


図3 医用画像診断支援システムにおける効果音の生成処理

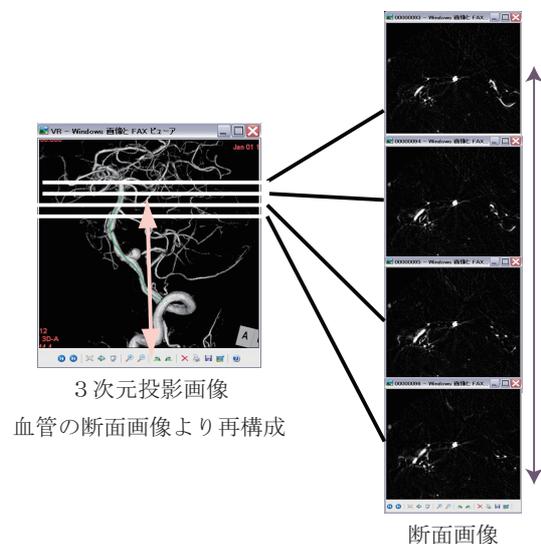


図4 断面画像（血管）

- (3) 効果音の割り当て
- (4) 診断用画像データの生成

画像の閲覧時に診断者は、基本的に先頭から順番にこれらの画像を見ていくが、必要に応じて、画像を前後させることがある。そのような場合でも効果音を出すことができるように、全ての画像に対して効果音を付与し、表示の際に対応する音を出力する。

3.2 画像特徴の抽出

3.2.1 画像特徴

効果音を作成するために、入力された画像から任意の特徴量を抽出する。この時に画像全てを対象とすると情報量が膨大になってしまうので、注視領域 (ROI, Region of Intension) を設定する。診断者は、まず矩形によって最初に注視領域を設定し、処理対象の領域を限定する。

画像に対しては、はじめに領域分割が行われる。その後、ROIにある各領域の特徴を計算し、図5に示すように各領域の特徴量として求める。画像の特徴については、現在、以下のような幾何特徴を抽出している。

- 面積
- 円形度
- 外接長方形面積
- 領域内平均輝度値
- 長径, 短径

これらの定量的に得られる特徴は、効果音における周波数や振幅などのパラメータとの対応付けが比較的容易である。また、各画像間の特徴量の変化は、音の変化として表現される。

しかしながら、例えば、図6に示す画像において、中央の比較的大きな領域(血管断面)の変化には気づきやすいが、で囲んだような領域においては、変化に気づきにくい傾向にある。また、断面画像の断層の解像度が非常に高い時には、画像間の変化があまりないために、連続画像の前後をみても違いがよく分からないことがある。

断面画像シーケンスに対する診断では、音の効果として連続画像間での領域面積や形状、平均輝度の微小な変化に対する効果音の生成が必要となる。これらは、特徴量が大きく変化する場合には、効果音がなくても画像の変化に気づきやすいが、画像間の変化が微小であるほど視覚的に確認することが困難である。

そこで、特徴量の変化に関して、それらを強調するために、差分画像を求め、そこから得られる特徴量を利用する。現在は、差分画像を2値化し、その面積などの幾何的特徴を利用している。

3.3 効果音の生成

効果音について、単純なサイン波に基づくピープ音を用いる。基本的には、以下の式によって表現される。

$$s(n) = A \sin\left(\frac{2\pi f_0 n}{f_s}\right) \quad (1)$$

この式のうち、振幅(A)と周波数($\frac{f_0}{f_s}$)に相当するパラメータを対応させる画像の特徴量に応じて変更することによって、特徴量に基づく多様な効果音を生成する。図7にパラメータの変化に応じたサイン波の変化を示す。

実際の利用においては、さらに変化を加えるために、ノイズとして別のサイン波を合成する。最終的には、以下の

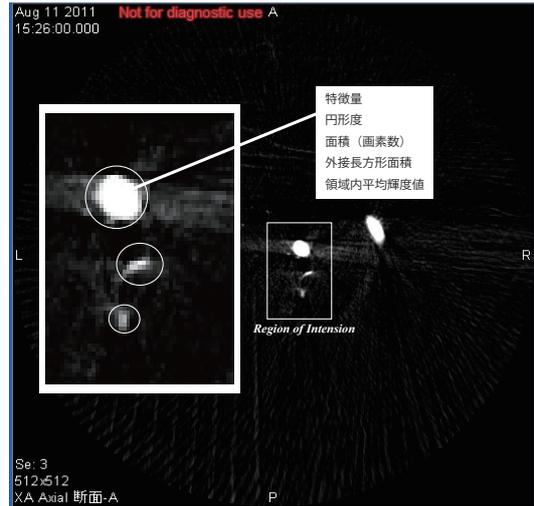


図5 画像内の特徴抽出



図6 断面画像(血管)の可聴化が難しい例

ような式で表現される。

$$s(n) = A' \sin\left(\frac{2\pi h_1 n}{f_s}\right) + A \sin\left(\frac{2\pi f_0 n}{f_s}\right) \quad (2)$$

3.2で得られた特徴量を用いて、ROIに含まれる領域のそれぞれの特徴量について対応する効果音を生成する。例えば、図5に関しては、一つの領域に一つの音が生成されるため、最終的に3種類の効果音が生成される。また、生成された音は継続的に提示されるため、対応する画像が表示されている間ずっと出力されている。

例えば、画像特徴とこれらのパラメータとを対応付けるを比例とすると、特徴量が大きければ、音量なども大きくなるというように設定される。

また、最終的に、各画像から得られた特徴量とその画像を中心とした差分画像から得られた特徴量を合わせて、効果音と対応付ける。

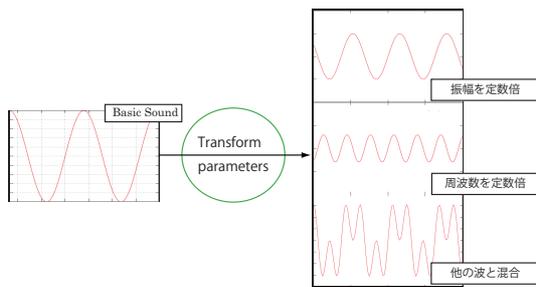


図 7 パラメータの変化によるサイン波の変化の例

3.4 効果音の割り当て

医用画像診断、特にスクリーニングを行っているときに、診断者は、異常箇所を発見するために、膨大な数の断面画像を調査する必要がある。この時に、診断者の興味症例や経過時間、患者の状態によって、画像内における異常箇所の見え方は大きく異なるため、画像特徴と効果音の対応付けを一概に決めることはできない。例えば、単純に特徴量と音量や音の高さを比例関係にした場合に、特徴値が非常に大きくなったときに非常に不快な音になったり、人間の可聴域を超えたりするなどの恐れがある。また、面積が大きい領域と小さい領域では、変化の割合が同じでも、変化量が異なるため、小さい領域の変化には気づきにくくなる。

したがって、画像特徴量と音のそれぞれに対して正規化を行う必要がある。また、あらかじめ全ての画像に対して特徴量を抽出し、その総変化量から最大値と最小値をあらかじめ求めることにより、このような問題を回避する。図 8 に示すように、画像特徴と差分画像特徴を併用して、全ての画像に効果音を付与する。最終的に、画像のシーケンスに従った効果音のシーケンスが得られる。

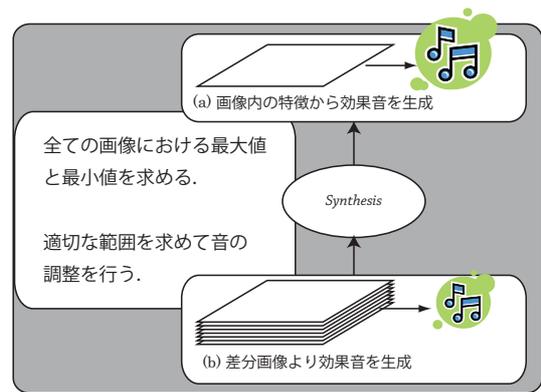
3.5 ユーザインタフェース

本手法の利用者インタフェースを図 9 に示す。インタフェース上の画像表示部には、適宜処理中の画像が表示される。診断時には、サイズを変えてできるだけ高解像度で閲覧ができるようになる。音に関しては、つまみなどを設定し、好きな音へ調整可能となるようにする。領域分割された画像とひとつ前の画像との差分画像が提示される。診断者は、マウスのホイールを用いて画像を一枚ずつ閲覧する。

4. 評価

4.1 評価ツール

可聴化による効果音付与を行う支援手法の有効性を評価するために、テストベッドとして図 4 に示す血管断面画像に対する診断支援ツールを構築した。画像で逐次的に血管断面トレースすると、例えば動脈瘤などがあると血管直径が突然大きくなったりする。ここでは、血管の画像特徴の



一枚の画像に対して効果音を割り当てることにより、最終的に効果音のシーケンスを構成する

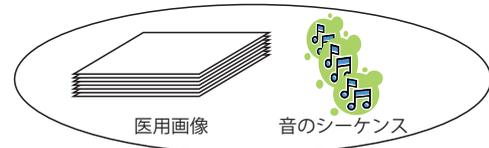


図 8 効果音の割り当て



図 9 ユーザインタフェース

状態を音で提示することにより、疑似的に動脈瘤を検出するタスクを行ってもらおう。

上記症例の画像については、大分大学医学部放射線科より提供していただいた画像から 2 種類（異常箇所なし，異常箇所あり）を選択し実験データとする。なお、それぞれ約 1000 枚からなるシーケンスとなっている。

図 10 に示すように、画像特徴に基づいた効果音を生成した。実際は、以下のような流れとなる。

(1) 画像特徴の抽出

濃淡画像として表現される画像に関して、ROI として画像中心から適当な画素数の矩形領域を設定し、最初に領域分割を行い、得られた各領域ごとに以下の特徴量を求める。

- 領域の面積（画素数）

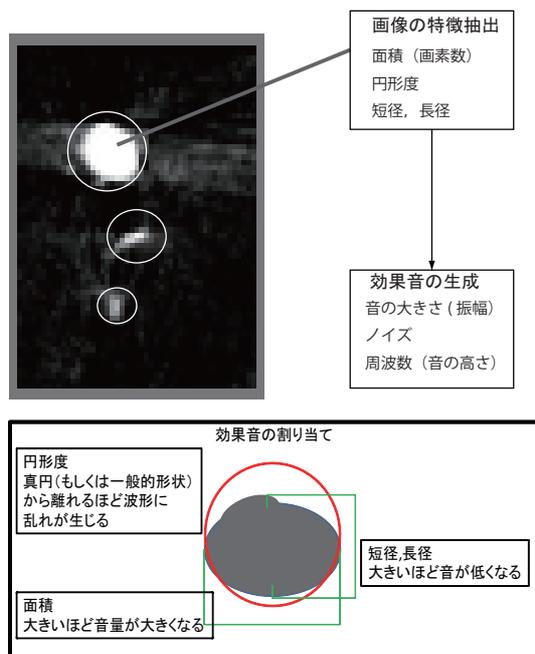


図 10 評価システムにおける効果音の生成

- 円形度
- 短径と長径

(2) 効果音の生成と割り当て

効果音については、上記の画像特徴に対して、図 10 のようにパラメータを対応付けた。効果音は、サイン波によるピーブ音を用いて生成し、その周波数と振幅を変化させた。また、円形度に関しては、他のサイン波と合成することにより、ノイズを加えた音を生成する。効果音の割り当てに関しては、現段階では、画像の特徴量と音情報を定義するパラメータを明示的に比例させることにより対応付ける。実際の効果音の聴こえ方は、図 10 下となる。

このような評価用テストベッドを 8 名の工学部学生に利用してもらい、簡単な評価実験を行った。評価実験用のタスクは以下の通りである。

- (1) 図 6 に示すような血管の連続断面画像を 3 件 1000 枚程度準備する。3 件の内訳は以下の通りである。
 - 対象画像なし。
 - 1 枚だけ対象画像あり。
 - 2 枚対象画像あり。
 被験者は、どの画像シーケンスにどれだけ対象画像が含まれているかは知らない。
- (2) 最初にこちらがあらかじめ決めておいた特定の画像を 3 枚提示する。ただし、血管が急に太くなるといった特徴的な画像を選択している。この合計 3000 枚の中から、3 枚の画像を探してもらう。
- (3) 効果音ありで画像を調べるグループとなしで調べるグループに分かれてもらい、提示した画像を探してもらった。

(4) 実験終了後に本手法に関するインタビューを行った。作業にかかる時間は、双方とも 3 分程度で大差はなかった。また、正答率 (正答数 / 回答数) は、音なしが 50 %、音ありが 62 % で多少精度が向上していた。この時、音ありグループに 1 名だけ誤答を繰り返す学生がいた。

実験後のインタビューでは、以下のような回答が多く得られている。

- 画像が変化した事が音でも分かった。
 - 画像を見るだけよりは長く続けられそうである。
- 一方で、以下のような意見も聞かれた。
- 画像を見ただけでも変化は分かる。
 - 音を聞き続けると集中しづらい。

このように、ネガティブな意見も聞かれたが、注意喚起を行うための効果音の効果に関しては、有効な結果が得られた。

4.2 検討

今回の評価では、簡単に一通り眺めてもらい、異常個所が見つかるかどうかを考えてもらうといった非常に単純な操作を行ったため、実際の運用形態とは必ずしも一致していない。できるだけ実際の環境とデータを再現したうえでの実験を行った上での評価を行う必要がある。また、画像診断を通常行う医師に被験者になり実際にシステムを運用してもらい、評価してもらわなければならない。その上で、システムに必要な機能の取捨選択をし、システムを洗練していく必要がある。

評価システムの実験を通して、検討事項として以下のことが挙げられる。

4.2.1 効果音の生成

定常的なピーブ音は長期間聴くと耳障りとなり、正常な判断の妨げとなることがあった。具体的には、作業の正答率が顕著によくなかった事例が生じた。そこで、以下の事項について検討する必要がある。

- 音色, 波形: 音色はこのようなインタフェースの構築に関して重要である。現在はサイン波を用いているが、状況に応じてパルス音やのこぎり波形のような音の利用も考えられる。また、MIDI などを利用して、聴き心地の良いピアノ等の楽器の音の利用も考慮する。そのために、音と画像に関する感性評価を行い、聴き続けても苦にならない音のデザインを検討する。
- 音量: 効果音による特徴の表現に関して音の大きさは重要な要素であるが、大きすぎたり、小さすぎたりすると、集中できなくなり、思考の妨げになる恐れがある。上記の音色とも合わせて適度で心地のよい音の生成を試みる。
- リズム, 鳴動時間: 今回の評価実験では、触れていないが、効果音を生成する上で、リズムは非常に重要で、また、情報をのせやすい要素である。それと同時に各

音の鳴動時間に関しても検討を行う必要がある。現在は、鳴り続けることを前提として実装を進めているが、画像間における音の変化が問題となるので、必ずしも常に鳴り続ける必要はない。そこで、適当に打ち切るような音の生成法を検討する。

4.2.2 画像と音との対応付け

現在の評価システムでは、比較的に見て分かりやすい画像特徴に基づいた効果音を生成し表現しているため、利用者にとっては冗長な情報の提示となっている。画像特徴が分かりにくいのであれば、この冗長性は、特徴の検出に有効に作用するが、分かりやすい特徴の時には、邪魔に感じてしまうことがある。画像特徴の見分けやすさを考慮に入れながら、画像特徴と効果音との関連付けを考慮する必要がある。

画像シーケンスの中で一枚の画像から異常が発見できる場合、医師は、その一枚を見つけるために集中して調べる必要がある。現在のツールで動脈瘤のように血管が途中で太くなったり、逆に細くなったりする点が異常箇所となる場合、画像特徴の急激な変化があった時に、医師は、異常箇所を発見しやすくなる。しかしながら、診断者は、異常があった画像を見つけることが最終目的ではなく、その前後の画像を見直して詳細に調査しなければならない。そのために、経験と専門知識をもって、画像シーケンス内に含まれるコンテキストを考える必要がある。

現在も差分特徴を利用しているが、前後の画像との差分情報を抽出し、それをもっと有効に効果音に反映する方法を検討している。この場合、画像には表れない前後の画像との関係も効果音として表現されるため、画像間の変化には敏感になると考えられる。しかしながら、現在、利用者が見ている画像に対応する効果音が生成されていないため、診断中に混乱してしまう恐れがある。また、画像による差分の計算に関する知識が要求される。

4.2.3 インタフェース

実際の診断においては、医師の意図に応じて効果音を生成する必要がある。この時に、例えば、図 11 に示すように、見つけたい画像特徴が単純ではなく、生成される音も複雑な規則を使わないといけない場合、画像や音声に関する知識の乏しい非専門家は、設定が困難なため、利用することができなくなってしまう。

なるべく簡単な GUI 操作によって医師の注視したい特徴に合わせた音情報の生成と出力を、画像や音声の非専門家である医師が指定しカスタマイズできる環境を構築する必要がある。対象とする画像と症例に応じて適切な特徴量が抽出されるように設定されなければならない。そこで、診断者が自由に獲得する特徴量を選択することのできる仕組みを構築すべきである。

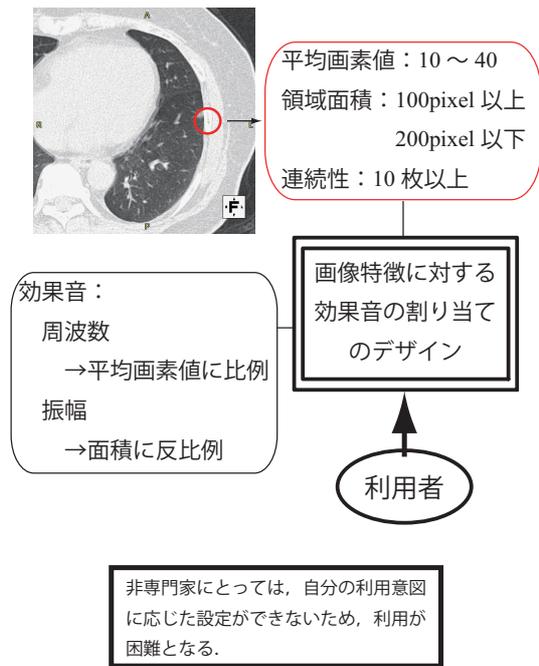


図 11 特徴と効果音の複雑な対応付けのデザイン

5. まとめ

本研究では、医用画像の画像特徴を可聴化し、その特徴に基づく効果音を用いて表現することで、医師の画像診断作業を支援する手法について提案した。血管断面を診断する評価ツールを構築し、その有効性に関して評価を行った。

今後は、実際に画像診断に携わる方を対象として、効果音に関する詳細な評価実験を行う。また、より詳細な設定が可能なシステムとインタフェースの構築を行っていく。また、例えば、監視カメラにおける効果音生成の有効利用など他の分野への適用とアプリケーションの枠組みについても検討していく。また、複数モダリティによる情報提示について、音の生成だけではなく、触覚による提示についても検討していく。

参考文献

- [1] K.Do, *Current Status and Future Potential of Computer Aided Diagnosis in Medical Imaging*, The British Journal of Radiology. Special Issue, pp.S3–pp.S19, 2005.
- [2] 藤田広志, 医用画像のためのコンピュータ支援診断システムの開発の現状と将来, 日本写真学会誌 2003 年 66 巻 5 号, pp.484–pp.490, 2003.
- [3] X.Huang and F.Zhang, *Mult-modal Medical Image Registration Based on Gradient of Mutual Information and Hybrid Genetic Algorithm*, Proc. of The 3rd International Symposium on Intelligent Technology and Security Informatics, pp.125–pp.128, 2010.
- [4] C.Doukas, I.Maglogiannis and K.Karpouzis, *Context-Aware Medical Content Adaptation through Semantic Representation and Rules Evaluation*, Proc. of the 3rd International Workshop on Semantic Media Adaption and Personalization, pp.128–pp.134, 2008.
- [5] R.Singh, M.Vatsa, A.Norrne, *Multimodal Medical Image*

- Fusion using Redundant Discrete Wavelet Transform*, Proc. of 7th International Conference on Advances in Pattern Recognition, pp.232–pp.235, 2009.
- [6] D.Sonntag and M.Moller, *A Multimodal Dialogue Mashup for Medical Image Semantics*, Proc. of the 15th international conference on Intelligent user interfaces, pp.381–pp.384, 2010.
- [7] A.Rodriguez, J.Labra, G.Hernandez, J.Gomez and R.Gomez, *ADONIS:Automated Diagnosis System based on Sound and Precise Logical Descriptions*, Proc. of 22nd IEEE International Symposium. 2009
- [8] 吉田翼, 木谷クリス真実, ペロンジーサージ, シュレイケビン, *EdgeSonic: 視覚障がい者のための画像特徴の可聴化*, 第 18 回インタラクティブシステムとソフトウェアに関するワークショップ WISS2010, 2010.
- [9] Peter B.L.Meijer, *An Experimental System for Auditory Image Representations*, IEEE Trans. on Biomedical Engineering, Vol.39, No.2, 1992.
- [10] Ivan Kopecek and Radek Oslejsek, *Hybrid Approach to Sonification of Color Image*. Third 2008 International Conference on Convergence and Hybrid Information Technology, pp.722–pp.727, 2008.
- [11] Jing Su, Alyssa Rosenzweig, Ashvin Goel, Eyal de Lara and Khai N. Truong, *Timbremap: Enabling the Visually-Impaired to Use Maps on Touch Enabled Devices*, International Conference on Mobile Human Computer Interface, pp.17–pp.26, 2010.
- [12] Jeremi Sudol, Orang Dialameh, Chuck Blanchard, Tim Dorsey, *LookTel-A Comprehensive Platform for Computer-Aided Visual Assistance*, IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp.73–pp.80, 2010.