C-09

Eyeke : カメラで聞くインタラクティブ音表現システムの提案

Eyeke : Interactive Sound Generation from Video Camera Images

奥中 健史 †　　外村 佳伸 †
Takeshi Okunaka　　Yoshinobu Tonomura

1. はじめに

近年，人とコンピュータとの関係を，ディスプレイ/マウス/キーボードだけでなく，我々の身の周りの環境を視野に入れたインタラクションシステムとしてとらえる研究が盛んに行われている．

人間は周りの環境から視聴覚や他の感覚器を介して情報を得ているが，意識上で意味につながる高次の認識のみならず，入力情報の特徴レベルの情報も無意識下の処理に大きな影響を与えている．そこで我々は，こうした特徴レベルの情報をインタラクションシステムに積極的に生かすことをめざしている．例えば，ユーザの身の回りの環境を映像情報として入力を行い，その映像情報の特徴レベルの情報を抽出し，その特徴ごとに自律的に様々な処理を施し，新たな表現としてユーザに提示する研究を進めている．

本論文では，その一環として，視覚情報としての映像入力に対し，その特徴情報を用いて聴覚情報としての音に変換するシステム **Eyeke** を提案するとともに，プロトタイプ化したシステムについて紹介する．

2. 画像情報から音への変換

画像情報を用いた音表現に関しては，視覚に障害がある人に音を用いて景色を伝えるシステムや，アートとして画像を音に変換するシステムなど，様々な目的や観点で行われている研究が古くからある．Cronly-Dillon らは，画像からエッジなどの特徴を抽出し，あらかじめ用意した対応する音を鳴らすことにより，視覚に障害がある人に形状を認識させる手法を提案した [1), 2)]．また，小林らは，音を用いてシーンの理解を行うため，頭部に固定したカメラ画像からランドマークの位置を検出し，3 次元音響装置を用いて，その位置から聞こえてくるような音を鳴らすことにより，視覚に障害のある人に歩行の誘導を行うものを提案した [3)]．楽器として画像情報から音表現を行っている研究の例として，Fels らのインタラクティブダンス楽器 Iamascope がある [4)]．このシステムは，ユーザの単純な動きを検出・画像処理し，画像を分割した各領域毎に音を出すかを判断することによりメロディー生成を行うものである．

本論文では，映像の特徴情報を生かすインタラクションシステムを目指す我々の研究の観点から，映像に現れる色に着目し聴覚情報へ変換するシステムについて述べる．本システムは，人間が関心を持つシーンや対象物を目で捉えるという行為に着目し，手持ち web カメラを用いて対象物をとらえることにより，音を発するものである．すなわち，映像中の色の具合をインタラクティブに音で聞くシステムである．また，応用の一つとして，インタラクティブな独特の楽器としての展開についても触れる．

† 龍谷大学大学院 理工学研究科， Graduate School of Science and Technology ， Ryukoku University

3. Eyeke

3.1 基本コンセプト

目に見えない温度を色に変換することで，人に直感的に温度を見せることが効果的なように，我々は特徴レベルの情報を処理・変換し，利用者に何らかの形で示すことにより，人の感覚特性を生かすインタラクションシステムを構築していきたいと考えている．

図 1 は，人間が視覚を通じて感覚・知覚からさらに高次の認識をする中で，視覚情報の特徴情報に処理(例えば強調，他メディアへの変換，現象の検出など)を加えることにより，感覚・知覚レベルで効果的に感じることができる何らかの表現に変換する処理の位置づけを示したものである．

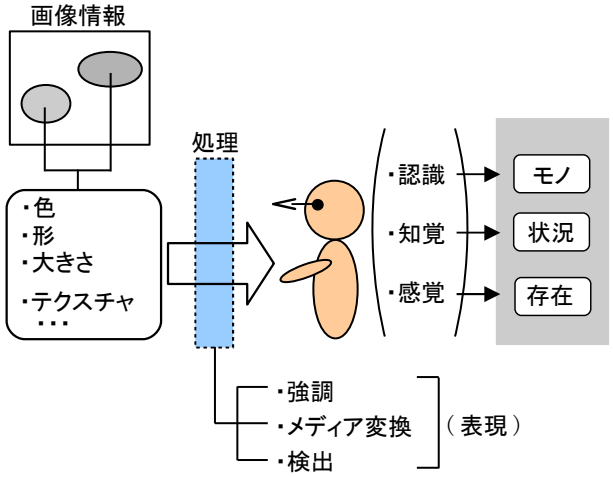

図 1 ：情報の処理･変換モデル

本システムは，関心のある対象にマイクでインタビューする様にカメラをかざし，その映像を分析し，特徴情報ごとに処理を行うという基本コンセプトから， Eyeke (Eye Mike)とした．今回具体的には，2 次元情報+α(時間軸)として入力される映像中の物の存在を，色情報を用いて処理し，1 次元情報+α(時間軸)としての音情報に変換することを試みた．

本論文では，視覚的な情報として，映像中のモノ（背景を含む）の存在を念頭に置いた．それらのモノは，形・色・大きさ・テクスチャなど様々な情報特徴を有しており，更にそれらの特徴は時間とともに変化するものもあれば，定常的なものもある．今回は，まずモノの定常的な存在に関わる映像中の情報特徴を大まかに捉えることとした．ここでは，モノの存在を連続検出される色で代表することにし，その位置と大きさ(面積)を基本的な一次特徴として用いることとした．

Eyeke では，その一次特徴を用いて音表現することにより，利用者に音でモノに対応する色を感じる体験を生もうとするものである．

3.2　映像情報から音表現へのマッピング

Eyeke では，入力された web カメラ映像から抽出される色情報をもとに音表現にマッピングを行う．

まず，一次元情報としての音への変換を行うため，同じように色を一次元情報として扱える色相値として扱うこととした．色の違いをある程度明確にするために，色相値を限定色にしたインデックス色（例えば 0 から 15）として表し，音情報についてもそれに対応して音楽上の離散的な限定音（音階）にマッピングすることとした．具体的には，音楽で用いる 12 平均律による中央のドから 1 オクターブ分の 12 音+1 音で音表現を行った．また，映像中の各色相値の量(面積)を音量に，位置を音の左右バランスに対応させることにより，映像中での色の存在感の程度と，大まかな画面内の位置が音表現からも判断できるようにした．映像と音がともに時間軸をもったメディアであるため，各色の面積の増減や画面内の動きなど，変化をリアルタイムで表現することが可能である．さらに，存在しているモノの形や分布を音色の変化に割り当てることで，より詳しい表現も可能となる．

4.　プロトタイプシステム

以上の方針に基づき，Eyeke のプロトタイプシステムを設計・実装した．

4.1　システム基本環境

Windows 7 を搭載した PC を使用し，映像の撮影は web カメラ(有効画素数：最大 200 万画素，1600×1200pixel，最大 30 フレーム毎秒)を用いて行った．

ソフトウェア開発は，Java によるソフトウェア開発で最もよく用いられる Eclipse 上でプログラミング言語 Java を用いて行った．また，Java 上でマルチメディアを扱うために，Java Media Framework 基本プラットフォームを用いた．音表現制御には音階表現が容易な MIDI を用いた．システムの構成を図 2 に示す．手持ち web カメラにより入力された映像を PC で処理・変換し音として出力する．

Eyeke の情報処理フローを図 3 に示す．まず，web カメラから入力された映像を分析し，後述する方法により限定色として色抽出を行う．そして，その限定色を色ごとに分析し，面積・位置の特徴量の抽出を行い，それらの特徴量を音の特徴量(音量・左右のスピーカーの出力割合)へと変換することで，色情報を音階へマッピングする．

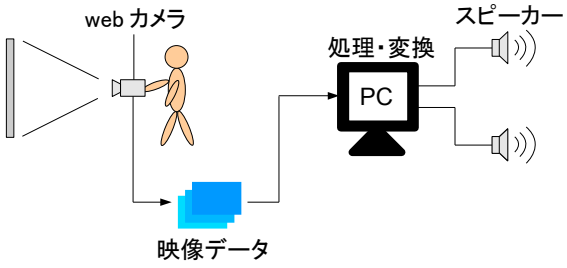

図 2 ： Eyeke のシステム構成図

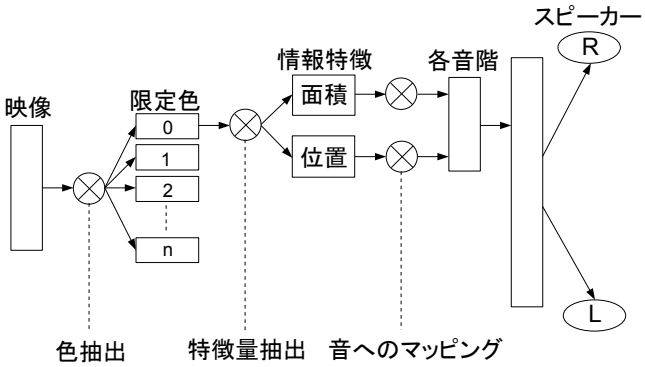

図 3 ： Eyeke の処理フロー

4.2　映像情報処理

Eyeke の特徴は手持ち web カメラにより，利用者が身の回りの環境を捉え，リアルタイムに映像情報から音表現出来ることである．したがって，実際の映像中の変化からの音表現への応答時間が速くなければいけない．また web カメラでインタラクティブに簡単に操作出来ることを目指すため，操作上の状況変化にロバストな色情報の抽出が必要となる．

4.2.1　限定色の決定手法

今回の提案手法における，各色相値から音階へのマッピングに当たっては，どの色相値をどの音に割り当てるかについて，処理上の何らかの基準が必要である．そこで，まず反射光ベースの色の基準として，色出力を計算によりコントロールできる紙へのカラープリントを用いて，その出力を本システムで撮影して色相値の抽出を行い，この検出値を元に，音への具体的なマッピングを決定することとした．しかし，環境としてのカメラ／プリンター／紙質／外光，さらにカメラと対象の位置関係や影などの周囲状況等，様々な要因があるため，常に計算上の色相値と同じ値の色相値が検出されるとは限らず，系の特性を知ることと，外乱による各色相値の変動の程度を知る事が必要となる．そこで，基準を作る上ではカメラ，プリンター，紙などについては，同じものを使用することで環境の安定を図った上で特性を調べ，外光など外的な動的変動要因については各色相値の変動幅を考えた上で，本システムで正確に判断できる色数に限定することで処理の安定化を図ることとした．色数の目標は音階 1 オクターブ分 12 音＋α である．その決定方法について次に述べる．

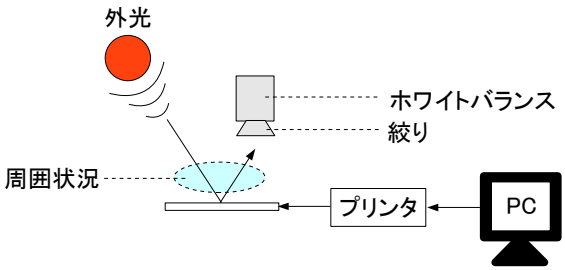

図 4 ：色識別の実験環境

検出される色相値には変動が伴うが，その要素には，印刷と撮影に伴う色相値の分散と色相値自身のシフトがある．予備検証で，まず検出色相は中心値±10°前後(色相：0～360°)内に収まる事が分かったが，問題は条件変動による中心値の振れである．その色相値の中心値に最も影響を与えるのは外光の変化と，それに対応させるカメラ機能としてのホワイトバランスである．外光は場所や光源によって変化は大きいものの，利用時点では，その状況として比較的安定するため，目的とする利用環境において，ホワイトバランスを調整して次に述べる基準特性にできるだけ近づけること，すなわち後述するキャリブレーションを行うこととした．

そこで，まず実験室環境で検出色相の特性を調べ，ホワイトバランス等のカメラ調節をし，その上である程度の色相値の分散による変動を考慮に入れて，分別できる限定色を検討した．最初に計算機により一定間隔毎の色相値を紙にカラープリントし，それを web カメラで撮影し，色相値の検出を行った(図 4)．図 5 はカラープリントによる反射光ベースの計算機出力色相値(以降，計算色相値)と検出された色相値(以降，検出色相値)との関係を示したグラフである．図 5 は横軸がカラープリントした計算色相値，縦軸が検出色相値の例である．色相値は 5°刻みで変化させている．色相値の変化に伴って，傾きが途中で大きく変化していることがわかる．このグラフの傾きが限定色化する上での分離可能な幅に大きく影響することになる．グラフ上で傾きが大きい程被写体となる色相値の変化に対し，検出色相値が変わる程度が大きく，逆に傾きが小さい程変化が小さい．すなわち，検出色相値の一定の変化を起こすための計算色相値の変化が，傾きが小さい範囲では大きい必要があり，傾きが大きい範囲では小さくて済む．計算色相値と検出色相値の関係が非線形のため，計算色相値上では等間隔に分割できず，図 5 の特性から区間を決定する必要がある．

次に色相値の分散を考慮するため，狙った検出色相値からどの程度値が振れるかを検討した．中心値からの分散幅を α とすると，検出可能な限定色の数は，(1)の式から求められる．後述するように，振れは最大で±12°(24°幅)であることから，可能限定色数は(1)に当てはめると 15 色となる．

$$\text{可能限定色数} = 360° / 2\alpha \qquad (1)$$

そこで検出色相値軸上で，24°刻みで区間をとり，その中心値を限定色の各色とすることとし，実際には図 5 の特性に合わせて実験的に調整し，図 5 の横線で示す 15 の色相値を決定した．

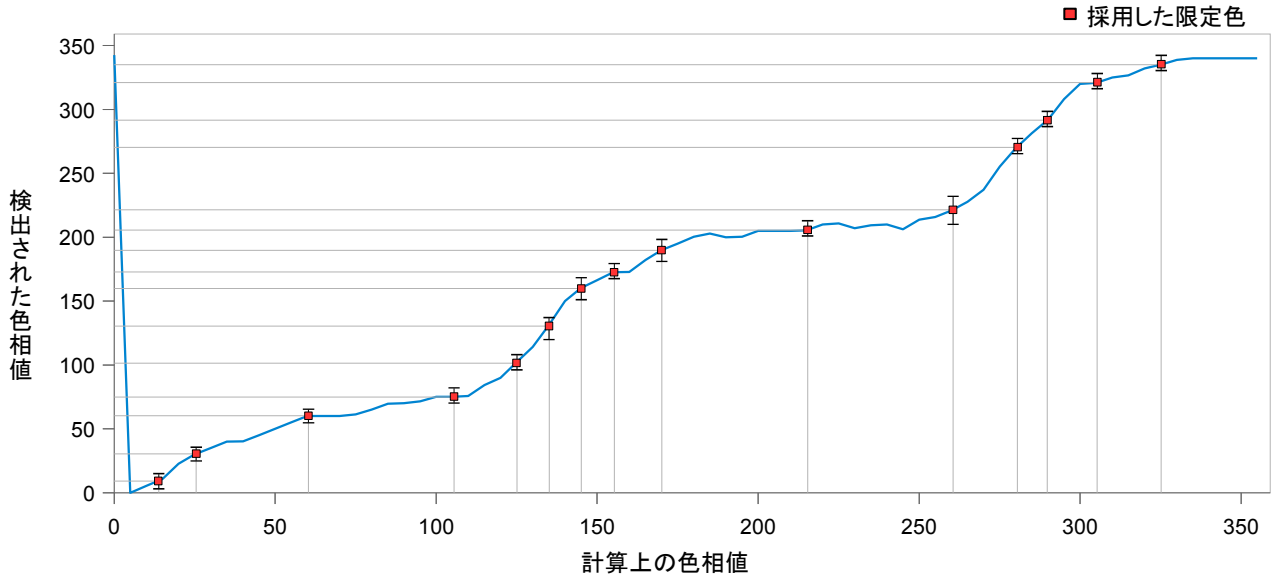

図 5 ：検出された色相値データ

表 1 ：各色相値と各音階の対応表

音階		ド	ド#	レ	レ#	ミ	ファ	ファ#	ソ	ソ#	ラ	ラ#	シ	ド
音名		C4	C4#	D4	D4#	E4	F4	F4#	G4	G4#	A4	A4#	B4	C5
周波数(平均律)		261.6	277.2	293.7	311.1	329.6	349.2	370	392	415.3	440	466.2	493.9	523.3
色相値		10	25	60	105	125	135	145	155	170	215	260	280	290

表 2 ：限定色を撮影した際に検出された色相値の分散と中心値との誤差

限定色	10°	25°	60°	105°	125°	135°	145°
推定分散	6°~19°	31°~39°	51°~59°	61°~74°	76°~89°	86°~104°	116°~134°
中心値との誤差	(-6°,+7°)	(-4°,+4°)	(-4°,+4°)	(-4°,+9°)	(-8°,+5°)	(-10°,+8°)	(-9°,+9°)

155°	170°	215°	260°	280°	290°	305°	325°
141°~154°	171°~189°	196°~209°	206°~229°	281°~294°	306°~319°	331°~344°	341°~354°
(-5°,+9°)	(-9°,+9°)	(-4°,+9°)	(-12°,+11°)	(-6°,+7°)	(-4°,+9°)	(-4°,+9)	(-5°,+8°)

4.2.2　カメラキャリブレーション

前述した，ホワイトバランスを中心とする色相シフトの問題に対し，カメラキャリブレーションを行うことで安定した色相値検出が行えるよう試みた．まず，計算色相値全体を写し，web カメラの絞りを調節することで，色相ヒストグラムが全体に均一（彩度のダイナミックレンジを広くとれるところ）になるようにし，固定する．そして，次にホワイトバランスを調節し，前述で用いた色の計算機出力紙により，図 5 中の何点かで出力色相が合うようにする．これにより，図 5 の特性に極めて近い条件となる．

4.2.3　色から音階へのマッピング

前述の 15 個の限定色と音階へのマッピングを表 1 に示す．とり得る限定色は 15 色であったが，現段階では 1 オクターブ＋1 音(ド)を表現するため 13 色を用いて音階へのマッピングを行った．

4.2.4　限定色値キャリブレーション

これまで，印刷と撮影に伴う色の分散と色相値の中心値シフトの問題に対し，限定色の決定とカメラキャリブレーションを行うことにより安定した色相値検出を可能にした．しかし，図 5 と表 2 から分かるように，各限定色によって外的環境による変動の影響を受けやすい色とそうでない色があり，各限定色によって分散が異なる．そのため，より安定して全ての限定色の検出色相を基準特性と合わせるために，カメラキャリブレーションによりある程度基準特性と検出される色相値を近づけた上で，後で述べる限定色値キャリブレーションにより限定色の検出色相値の中心値を調整・決定することとした．

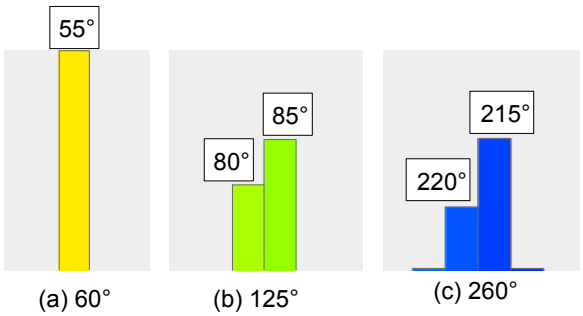

図 6 :限定色の 60°，125°，260° における
検出された色相ヒストグラム

本システムは，前述した検出された限定色の中心値とマージンの決定処理をコンピュータ上で行う．次にその処理方法について説明する．図 6 は限定色である(a) 60°，(b) 125°，(c) 260°の色相値を撮影した時の検出される 5°ずつ刻みの色相ヒストグラムである．全ての限定色を撮影し，検出される色相値の分散を調べた結果，(a) 60°で最も分散が小さく，(c) 260°で最も分散が大きいことが分かった(表 2)．今回，5°ずつ刻みで色相値検出を行ったため，(a)(b)(c)はそれぞれ 51°~59°，76°~89°，206°~229°の間で分散していると考えられる．このことを踏まえて，限定色を一つ一つ撮影し中心値の決定を行う．(a)のようにほぼ一点で色相値が検出された場合はその点を中心値(55°)とする．しかし，(b)(c)では分散が広く，計算により中心値を決定する必要がある．分散が広く見られる場合，検出される色相ヒストグラムは(c)のように山のような形状となる．よって，中心値は山の頂点，つまり最も多く色相値が検出された点となる．このことより，検出された色相値の量が多い上位 2 つの間に中心値が存在していると考えられる．具体的には，(c)の場合だと 215°と 220°の間である．中心値は，上位 2 つの色相値のうち，検出された量が多い方に近いと考えるのが妥当である．そこで，上位 2 つの色相値の比を求め，その比率から中心値の位置を決定した．具体的に(c)の場合で計算を行うと，215°では検出された量が[39]，220°では[84]であり，その比は 31.7：68.3 である．つまり，約 3：7 という比率となり，中心値は 218°にあると推測できる．

以上の処理により限定色全ての検出された色相値の中心値を決定し，中心値間の中点を求めることにより限定色の検出色相値区間を決定した．

以上述べてきた限定色の決定，カメラキャリブレーション，限定色値キャリブレーションという 3 つの手法を用いることにより，外的環境により左右されにくい安定した動作が可能となった．また，二つのキャリブレーションに要する時間は 1，2 分ぐらいであり，利用者の大きな負担にはならないと考える．

4.2.5　白・黒の扱いについて

黒と白については，色相上では扱えず，輝度と彩度により扱うことになる．白は何らかの照明下では，カメラで撮影した場合，てかり部(彩度・輝度大)となって検出され，動作上の外乱となることが多い．そこで，このような部分は色検出を行わずカットすることとした．また，黒については，彩度・輝度ともにほぼ 0 近いため，存在のないものとして扱うこととした．以上の考えより，白と黒については音階へのマッピングは行わず，検出しても

音出力を行わない無音とすることとした．このことにより，後述する本システムのインタラクティブな楽器への利用の場合では，音と音の切れ目を表現でき，背景として白と黒を積極的に用いることが可能となった．

4.3 Eyeke の展開

Eyeke は色に着目して映像を音に変換する基本環境としてのシステムだが，楽器，景色分析など様々な展開が考えられる．

より限定したインタラクションシステムとしての応用に楽器としての展開が考えられる．例えば，音楽の音譜のように，色による音譜いわば「色譜」を用いて，音楽を奏でることが出来る．図 7 は，ピアノのキーボードのように音譜対応の色を並べたものであるが，特定の楽曲に対応する色譜も数多く試作した．Eyeke は，映像中のモノの存在を音に変えるという基本コンセプトから，映像中に複数のモノの存在を検出した場合，同時に複数の音を奏でることも出来る．そこでカメラ画角内に複数色が入るようにすれば和音の演奏も可能である．また，曲の繰り返し部分などは，同じ部分を繰り返し映すことで簡単に演奏が可能である．また，Eyeke は色の面積と音量を対応付けしており，対象となるモノとカメラとの距離により音量を調節することが可能である．そのため，実際に演奏した場合，まるで指揮棒のような操作で距離による音量の強弱を付けることもできる(図 8)．色譜を筒状あるいは円盤状にし，回転させることによるオルゴールも作れる(図 9)．一方，紙媒体を映すのでなくコンピュータスクリーンにプログラムした色を映すことで，スクリーンにカメラをかざせば，音楽が聞こえるシステムにもなる(図 10)．図 10 では，ノートパソコンのスクリーン上に 2 色分の表示エリアがあり，時間を追って変化する様子を左の web カメラで撮影している．和音や輪唱が実現できた．

Eyeke の展開として景色分析への利用も考えられる．本論文で述べてきた，色の存在の音変換の他に例えば，映像中の変化のみを検出するフィルタを実装することにより，物体の動きを音表現できるのではないかと考えている．これは，映像と音という二つのメディアがともに時間軸を持っているため可能であり，映像の変化を音の変化として表現するものである．

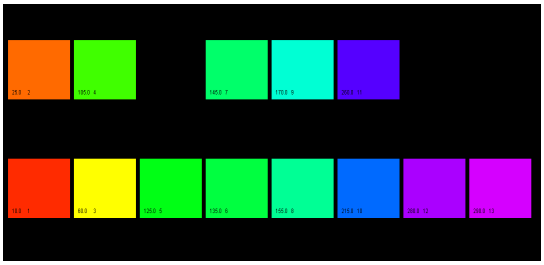

図 7 ：色譜の例（1 オクターブ分のキーボード型）

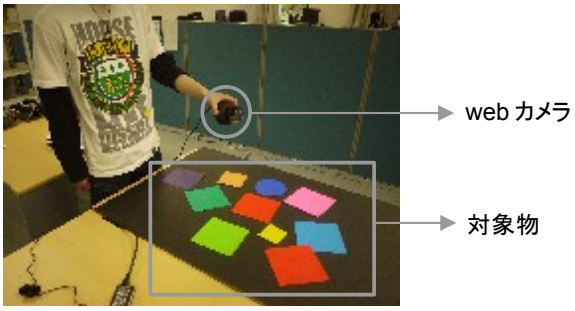

図 8 ： Eyeke の利用風景

図 9 ：色譜の例（円盤状のオルゴール型）

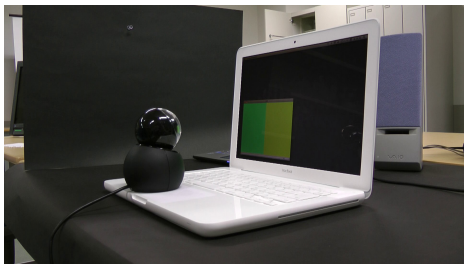

図 10 ：色のスクリーン表示による音楽演奏風景

5. おわりに

本報告では，特徴レベルの情報をインタラクションシステムに生かす研究の一端として，映像の特徴情報を用いて聴覚情報としての音に変換するインタフェース Eyeke を提案し，プロトタイプシステムを実装した結果について述べた．本システムは手持ち web カメラを用いることにより，利用者が自ら捉えたいものに対して，映像の特徴情報を音で表現することを可能にした．また，実験による限定色の決定とカメラキャリブレーション，限定色値キャリブレーションにより，外的環境に左右されにくい安定した動作が可能となった．

今回は，各色と各音階の対応を単純に行ったが，本システムは利用者の意識上で感じる特徴を強調したり，無意識刺激の提供を行うことで，利用者の気づきや感情に作用することを目的としており，色と音から受ける印象を形容詞表現を介して関連性を考えた上でマッピングを行っていく必要がある [5), 6)]．そのため，今後どの様な表現がどういう効果を生むかについても検討していきたい．また，Eyeke の今後の展開として，景色分析を行う際には，景色中に無数の色が存在するため，表現の幅を広げる必要があり，限定色数の増加を検討する必要がある．

文献

[1] J. Cronly-Dillon，K. Persaud，and R. P. F. Gregory：“The Perception of visual images encoded in musical form: a study in cross-modality information transfer”，*Proceedings. Biological science / the Royal Society*，Vol 266，pp. 2427－2433，(2000)

[2] J. Cronly-Dillon，K. C. Persaud，and R. Blore：“Blind subjects construct conscious mental images of visual scenes encoded in musical form”，*Proceedings. Biological science / the Royal Society*，Vol 267，pp. 2231－2238，(2000)

[3] 小林真，太田道男：“全方位センサと 3 次元音響を利用した視覚障害者用歩行誘導システム”，バイオメカニズム学会誌，Vol. 24，No. 2，pp.123－125，(2000)

[4] 間瀬健二，シドニー・フェルス，ダーク・ライナース：“Iamascope(インタラクティブ万華鏡): グラフィックな楽器の提案”，Visual Computing グラフィックスと CAD 合同シンポジウム，pp.91－96，(1998)

[5] 長田典子，岩井大輔，津田学，和氣早苗，井口征史：“音と色のノンバーバルマッピング―色調保持者のマッピングルール抽出とその応用，信学論，Vol. J86-A，No. 11，pp. 1219-1230，(2003)

[6] 井口征史他：“感性情報処理”，信学会ヒューマンコミュニケーション工学シリーズ/電子情報通信学会編，オーム社，(1994)