Coloresia: An Interactive Colour Perception Device for the Visually Impaired

Abel Gonzalez, Robert Benavente, Olivier Penacchio,
Javier Vazquez-Corral, Maria Vanrell, and C. Alejandro Parraga

Abstract. A significative percentage of the human population suffer from impairments in their capacity to distinguish or even see colours. For them, everyday tasks like navigating through a train or metro network map becomes demanding. We present a novel technique for extracting colour information from everyday natural stimuli and presenting it to visually impaired users as pleasant, non-invasive sound. This technique was implemented inside a Personal Digital Assistant (PDA) portable device. In this implementation, colour information is extracted from the input image and categorised according to how human observers segment the colour space. This information is subsequently converted into sound and sent to the user via speakers or headphones. In the original implementation, it is possible for the user to send its feedback to reconfigure the system, however several features such as these were not implemented because the current technology is limited. We are confident that the full implementation will be possible in the near future as PDA technology improves.

1 Introduction

Colour is an important feature of everyday life. Although highly saturated objects are not abundant in nature, we build and paint objects with highly saturated colours in an attempt to grab each other's attention, please each other, and transmit information. In the natural environment, colour helps organising scenes (blue is predominant in the sky, green in chlorophyll, brown in earth, grey in rocks, etc.) and crucially, it aids important survival tasks such as finding ripe fruit and leaves, detecting poisonous animals, and breaking luminance camouflage. In cities, colour highlights

Abel Gonzalez · Robert Benavente · Olivier Penacchio · Javier Vazquez-Corral · Maria Vanrell · C. Alejandro Parraga
Computer Vision Center / Computer Science Dept.,
Universitat Autònoma de Barcelona, Building O, Campus UAB, 08193 Bellaterra, Spain
e-mail: agonzgarc@gmail.com,
{robert,penacchio,jvazquez,maria,aparraga}@cvc.uab.cat

or simplifies important information (red for danger or stop, green for way-out or go, fast identification of known products, understanding of train/metro maps, etc.) and this fact has been exploited to such degree that we are surrounded by advertising, fashion, traffic signalling, etc. that relies on colour to transmit distinctive visual information. However, colour processing is not an easy feat: years of research and technology development have shown that to extract reliable colour and texture information in lexical form from natural images is far from trivial. The main problems to be addressed are not related to the technology available (medium to high-quality colour portable digital cameras are ubiquitous nowadays) but instead are related to the way humans sample and perceive the wavelength distributions of visible light. The human visual system has several mechanisms to extract meaningful information from the light that reaches the eye, filtering out the less important, more redundant patterns. These include a bias towards representing the reflecting characteristics of objects rather than the chromatic content of the illumination (colour constancy) [32], a tendency to enhance or suppress the perceived richness (saturation) of a colour according to the variability of its extended surrounds [2], and several other mechanisms which alter the perceived hue of an object according to its immediate surroundings (chromatic induction) [3]. On top of this, there are various complex cultural issues that affect the way we transmit to others the information about what we perceive (language). For example, not everybody agrees on which semantic labels to assign to the same wavelength signal, and everybody is familiar with the experience of arguing about the colour of a piece of clothing or a newly painted wall. However, anthropologists have found a set of 11 basic colour terms that are common to most evolved cultures (white, black, red, green, blue, yellow, grey, brown, orange, pink, purple) [4] which are a good starting point to model the universal attributes of colour naming.

1.1 Colour Vision and Colour Visual Deficiencies

Colour is everywhere, and its very ubiquitousness and vividness makes us forget that it does not exists in the world ”per se” but it is constructed by our brains from a few highly specialised neurons in our retinas. The delicate equilibrium of this neural construction becomes apparent when something goes wrong and our perception of the world becomes impaired. There are many forms of visual chromatic handicap, but some of the most common are impairments linked to deficiencies (or loss) of a given retinal photoreceptor. According to statistics compiled by the American Academy of Ophthalmology “*red-green colour vision defects are the most common form of colour vision deficiency. Approximately 8% of men and 0.5% of women among populations with Northern European ancestry have red-green colour defects. The incidence of this condition is lower in almost all other populations studied*” [5]. The rate of incidence of blue-yellow colour vision defects is the same for males and females (fewer than 1 in 10,000 people worldwide). Complete achromatopsia (a rare type of impairment where subjects do not see colours and only perceive shades of grey) affects an estimated 1 in 30,000 people. People with achromatopsia almost always have additional problems with vision including reduced visual acuity, increased sensitivity to light (photophobia), etc. When visual acuity impairments are higher than 20/200 (10% of normal vision in Spain) or the visual field is less than 20 degrees in diameter, sufferers are considered legally blind. In the U.S., there are more than one million legally blind people aged 40 or older (0.3% of the population) and only 10% of those are totally blind [5] (see Figure 1).

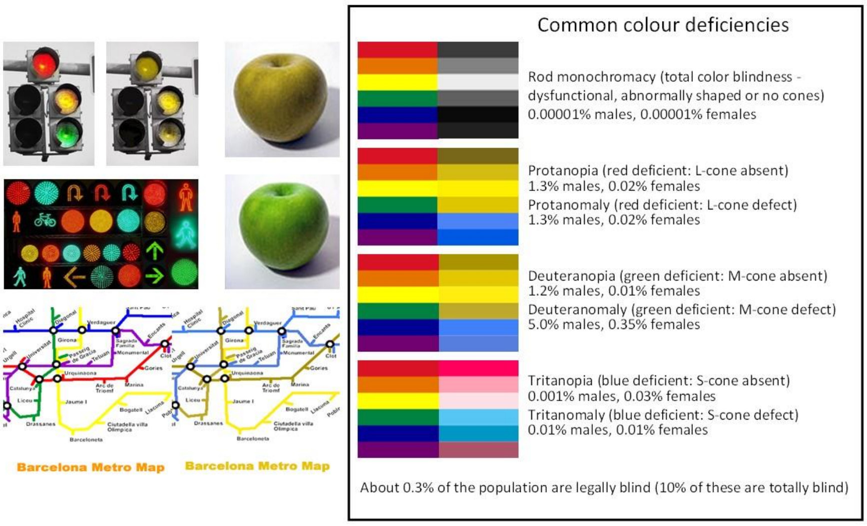

Barcelona Metro Map
Barcelona Metro Map
Common colour deficiencies
Rod monochromacy (total color blindness - dysfunctional, abnormally shaped or no cones)
0.00001% males, 0.00001% females
Protanopia (red deficient: L-cone absent)
1.3% males, 0.02% females
Protanomaly (red deficient: L-cone defect)
1.3% males, 0.02% females
Deuteranopia (green deficient: M-cone absent)
1.2% males, 0.01% females
Deuteranomaly (green deficient: M-cone defect)
5.0% males, 0.35% females
Tritanopia (blue deficient: S-cone absent)
0.001% males, 0.03% females
Tritanomaly (blue deficient: S-cone defect)
0.01% males, 0.01% females
About 0.3% of the population are legally blind (10% of these are totally blind)

Fig. 1 Dichromats (People with impaired colour vision) find it difficult to perform basic tasks that involve detection and semantic labelling of different colours. These range from detecting danger signals at pedestrian crossings, discrimination of ripeness in fruit, discrimination of colour-coded train and tube lines in maps, etc. The right panel introduces some statistics about common deficiencies and their prevalence in the U.S. population [5]. (See color version of the figure at: http://www.cic.uab.cat/Publications/)

Visually impaired people face a number of everyday problems, ranging from the mild to the severe. In particular, they may experience problems recognising different bi-colour or tri-colour Light Emitting Diodes (LED) traffic lights and in some physical arrangements, the position may not be a cue to their colours, as in the case of horizontal traffic lights. There is also the inconvenience of not being able to navigate the coloured maps of motorways, trains and tube lines, either printed on paper or on electronic media. Dichromats also complain that other people ”think that their choice of colours is strange” and that they cannot tell whether a piece of meat is raw or well done, or if a fruit is mature among other everyday problems.

1.2 Perceptual Interaction between Colour and Sound

Hearing is arguably the second most important way by which humans sense information about the world, and consequently sound is another important feature of everyday life. As with colour, we use sound to capture each other's attention, transmit information, and please each other.

Although they are processed by mainly separate neural mechanisms (and therefore studied by different disciplines), there is evidence that the mammalian visual and auditory systems may have many areas of overlapping. For instance, both systems share the ability to determine the speed and direction of a moving object, and to produce a unified percept of movement. Therefore, both types of sensory information have to merge or coordinate at some point. In addition, both systems have to coordinate and interact to direct attention to one modality or the other to control subsequent action [20]. More evidence of this neural mechanism overlap is provided by the involuntary cross-activation of the senses that occurs for a handful of individuals, in sound-colour synaesthesia, where auditory sensations spontaneously elicit visual experience. For example, when a key is struck on a piano a sound-colour synaesthete experiments a vivid colour sensation (see [42]) and this sensation may be different if another key is struck. However, if the same note is played the sensation elicited is internally very consistent over time. Many musicians experience this phenomenon [41].

Although individuals with sound-colour synaesthesia differ in their cross-modal associations, the sound-to-colour mapping they experience is not necessarily arbitrary. For example, the vast majority of them associate high pitch with light colour [42]. In addition, both non-synaesthete and synaesthete people share the same heuristics for matching colour and sound. The difference is that the cross-modal sensation is elicited involuntary for synaesthetes, whereas it involves a conscious initiative/effort for non-synaesthetes. All in all, it seems that sound-colour synaesthesia uses some common mechanisms of cross-modal perceptual interaction [42]. Accordingly, sound-colour cross-modal perception by synaesthetes is of interest for defining a colour-to-sound correspondence because it seems not to recruit privileged pathways between auditory and visual modalities.

Indeed extreme cases of synaesthesia are rare, however researchers studying how the brain combines information from different sensory modalities (i.e. cross-modal perception and multisensory integration) hypothesise whether it might be the case that all humans are synaesthetes to some degree and whether these naturally biased correspondences may influence the development of language [19].

Synaesthetic individuals seldom complain about their condition, and in many cases they claim that their lives have been enhanced by this ability to relate colour to sound or haptic information. This apparent "enhancement" has motivated us to apply current multimodal interactive techniques to deliver the information that is missing in one sense (vision) as a pleasant stream to other sense (hearing). In other words, we created a portable device (Android platform) that extracts semantic colour information from images in a manner compatible with the human visual system and conveys this information as a pleasant stream of music which does not overwhelm or bother the user (see figure 2). We also wanted to make the device "interactive", i.e. capable of receiving input from the user and "adaptive", i.e. capable of learning from the user input to improve its inherent properties. Unfortunately, some of the work towards this aim was not implemented in the prototype due to current limitations of the portable device technology. However, we are confident that at the current rate of technological improvement suitable devices will be available in the near future.

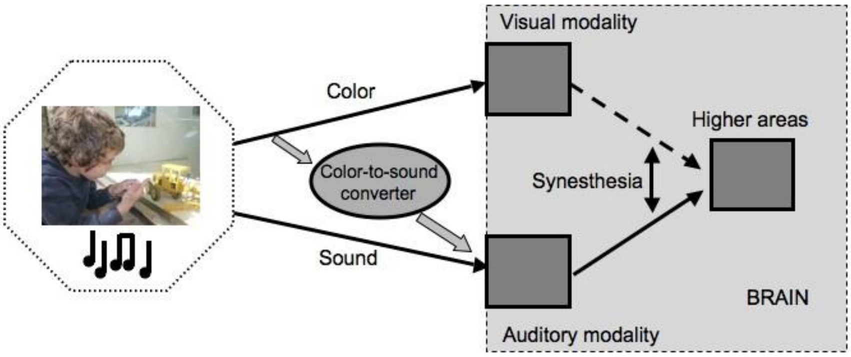


Fig. 2 Similarly to what happens in synaesthesia, the developed device converts colour information to sound

2 State of the Art

Until recently, conversion from colour to sound and vice versa had received more attention from visual arts than from science. Several techniques aiming to convert sounds or music to a visual presentation are included in what is known as visual music [24]. Although these first approaches did not exactly transform colour into sound, they were a first step towards the goal of expressing colour as music (see [21] for an historical account of colour-to-sound correspondences). In the last years, the idea of implementing aid devices for helping blind and visually impaired people to perceive colour through the representation of colour as music has received an increasing attention.

Cronly-Dillon *et al.* [25] showed the viability of representing some features from an image using music to describe its content to blind people. Their method selected different features of an image and represented each of them with a sound. The sounds for each part of the image were combined as a polyphonic melody that encoded the basic content of the image. Their experiments showed that blind people were able to interpret some images by hearing their associated melodies.

Following a similar line, Bologna *et al.* [26] proposed a method to transform coloured pixels into musical notes in order to describe image content for blind users. To this end, hue was divided in several sectors and was represented by timbre (see below), saturation was divided in four levels and was represented with different notes, and luminosity was represented by bass for dark colours and a singing voice for bright colours. Using this transform, the input image was segmented and the sounds corresponding to the colours of the main parts of the image were reproduced. Bologna *et al.* also proposed to use saliency detection techniques to focus the description on the most salient parts of the image.

A similar idea was proposed by Rossi *et al.* [27], who developed a prototype of a device that transformed colours into melodies. The system was developed as a game for children and was implemented in a portable bracelet with a small camera installed on a pointer that allowed users to select any point of the scene. The system was able to identify six colours (red, green, blue, yellow, purple, and orange) by dividing the hue circle of the HSV colour space in six sectors. Each of these colours was assigned to a musical instrument that played a melody that could be chosen from a set of five melodies. Additionally, for each colour, three to five divisions were set on the value dimension, and each of these subdivisions was identified by a different tone. Black and white were also considered as additional cases on this system. As in the approach of Bologna *et al.* the initial identification of colour names was not perceptual and this fact might be a drawback of both systems.

The approach which is closer to our purpose is the one by the visual artist and composer Neil Harbisson [22]. Harbisson suffers from achromatopsia, a visual condition that allows him to only see the world in shades of grey. To overcome his lack of colour perception, he designed a device called Eyeborg, which consists of a sensor that he wears on his head and points towards the direction he is looking at. Using a chip fixed to the back of his neck, the frequencies of light are converted into audible frequencies, which he interprets as a colour scale. Harbisson has developed two different conversion algorithms. The first one directly transforms seven light frequency ranges into seven sound frequencies. His second approach, divides the light frequency scale in 12 ranges corresponding to different colours and converts them in 12 musical notes. Both methods result in unpleasant and even heady sounds.

As we stated in the introduction, our goal is to develop a personal assistant implemented on a mobile device running under the Android platform. Several applications that acquire images with the device camera and are related to colour detection and identification can be found at the online shop for the Android platform, *Google Play* [23]. Some examples are *'This Color What Color?'*, *'Color Detector'*, *'Color Picker'*, and *'Color Blend'*. Although some of them give the name of the colours using synthesised voice, to the best of our knowledge, there is no application implementing a colour-to-sound transform algorithm to specifically aid visually impaired people.

3 From Colour Signal to Sound

We have built a prototype for colour name extraction that is able to, given a digital image, provide a list of the main colours of the objects present in it, in a manner consistent with the behaviour of human observers (see prototype schematics in Figure 3).

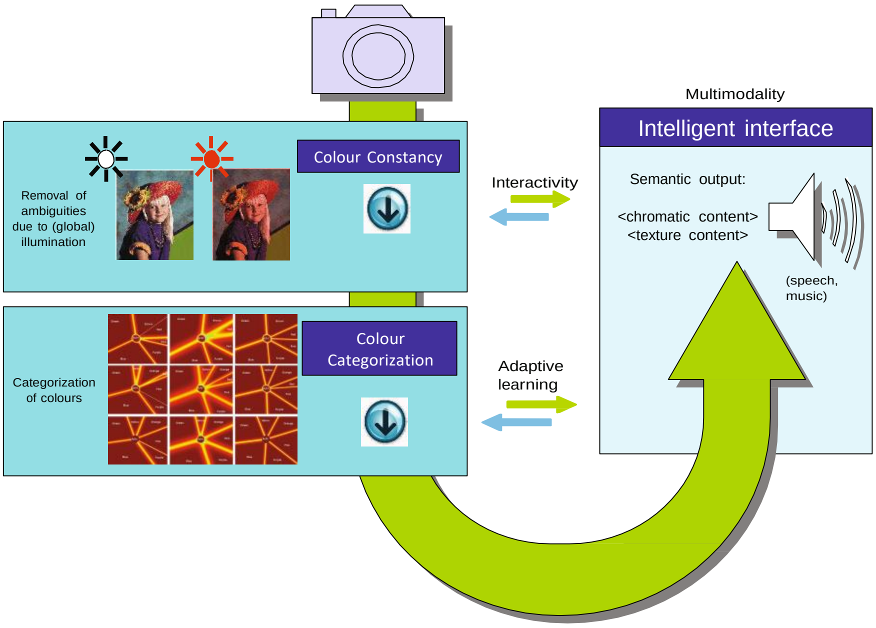


Fig. 3 Schematics of the prototype (See color version of the figure at: http://www.cic.uab.cat/Publications/)

The prototype is able to communicate this information to a visually impaired user in two modalities: words and music. The definition of visually impaired here ranges from dichromats to low vision or even blind users. In other words, we have built a portable system that acts on the output of a digital camera and reproduces the basic mechanisms that a human observer employs to identify the names of the colours of the objects present in the scene. The colour names are communicated to the user by means of synthesised music or alternatively, an automated voice system. We achieved this aim by:

- developing a human-based colour perception model to account for changes in perceived chromatic characteristics of the illuminant.
- developing a set of image descriptors to identify and label the main colours in images, in a manner similar to human observers.
- developing an interface based on natural language that is able to handle colour names.
- developing an interface based on sound that is capable to convert colour names into music

Our prototype was conceived as a portable device, based on a state-of-the-art personal digital assistant (PDA) with an embedded digital camera. Such devices are relatively inexpensive and provide the necessary capabilities to develop a software-based model that uses the digital camera (input device) as a first stage and delivers its results through the sound system (speakers/headphones). They have also an adequate user interface hardware (touch screen) for entering the necessary user corrections to improve the colour-naming algorithm. Figure 4 provides the schematics of the prototype design. The input data comes via the PDA camera's uncalibrated camera and the system applies an illumination removal algorithm to produce an image free of the colouring imposed by the illuminant. We use this representation to classify the content of the scene according to its colour names. The output of this algorithm comes in two alternative forms: as a voice through a voice synthesiser/speaker combination or as music.

In the following sections we explain in more detail the physical and perceptual properties of colour and sound that we are about to simulate and manipulate to achieve the "sonification" of the image, i.e. the transfer of colour information to the auditive system.

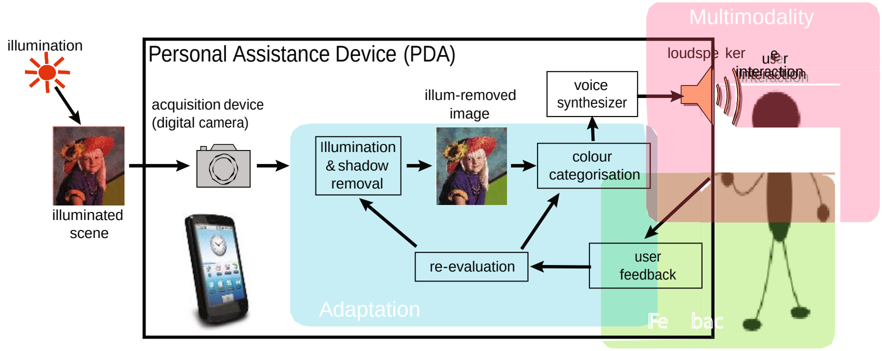


Fig. 4 Feedback, multimodality and adaptation and their role in the prototype (See color version of the figure at: http://www.cic.uab.cat/Publications/)

3.1 *Properties of Colour*

The wavelength content of the electromagnetic radiation that reaches our eyes is sampled in the retina by specialised neurons (cones), converted into neural information and transferred throughout different stages in the visual pathway. In the latest stages, the information is categorised. Categorisation is the process by which objects are differentiated and grouped, softening differences and favouring similarities among them, reducing an extremely complex world into cognitively tractable proportions. This reduction is extremely evident in the colour domain: from the nearly 2 million colours that can be distinguished perceptually we recover only about 30 colour categories which can be named by average subjects [7]. Although many colours can be distinguished and named, there is a group of 11 colour categories that are common to all advanced languages. They were defined by Berlin and Kay in their seminal work [4] and are thought to be inherent of the human neural machinery of colour categorisation [16, 17, 18]. These are black, white, red, green, yellow, blue, brown, purple, pink, orange, and grey, and they appear in a language in this particular order as the language becomes more complex. More complex languages tend to have more categories, but these are the most primitive.

To model this categorisation process as accurately as possible is a goal of many disciplines, from colour image reproduction to computer vision. Recent computational models of colour space segmentation are based on either natural scene statistics [8] or psychophysical data [9, 10, 11, 12, 13, 14, 15]. We implemented a colour space segmentation model on the model of Benavente *et al.* [11] because it has several advantages over others: it is implemented in CIELab colour space (a perceptually uniform space that has its lightness dimension built from relative luminance) and is parametric, i.e. can be easily adjusted depending on the user feedback. The model is built from fuzzy sets segmenting CIELab space in 11 regions and in its current implementation, it assigns to each pixel $\mathbf{p} = (L, a, b)^T$ a membership value between 0 and 1 to each colour category. Hence, for each pixel $\mathbf{p}$, a 11-dimensional colour descriptor $CD(\mathbf{p})$ is defined as

$$CD(\mathbf{p}) = \left[\mu_{C_1}(\mathbf{p}), \ldots, \mu_{C_{11}}(\mathbf{p})\right] \tag{1}$$

where each component of this 11-dimensional vector describes the membership of $\mathbf{p}$ to a specific color category and the component with highest membership value determines to which category the pixel belongs.

The value of each of the components of the colour descriptor is obtained from a triple-sigmoid with elliptical center (TSE) function given by

$$TSE(\mathbf{p}, \theta) = DS(\mathbf{p}, \mathbf{T}, \theta_{DS})ES(\mathbf{p}, \mathbf{T}, \theta_{ES}), \tag{2}$$

where ES is is an elliptical-sigmoid function which models the central achromatic region and is defined as

$$ES(\mathbf{p}, \mathbf{T}, \theta_{ES}) = \frac{1}{1 + \exp\left[-\beta_e\left[\left(\frac{\mathbf{u_1 R_\varphi Tp}}{e_x}\right)^2 + \left(\frac{\mathbf{u_2 R_\varphi Tp}}{e_y}\right)^2 - 1\right]\right]}, \tag{3}$$

and DS is a double-sigmoid function defined as the product of two oriented 2D-sigmoids given by

$$DS(\mathbf{p}, \mathbf{T}, \theta_{DS}) = S_1(\mathbf{p}, \mathbf{T}, \alpha_y, \beta_y)S_2(\mathbf{p}, \mathbf{T}, \alpha_x, \beta_x) \tag{4}$$

$$S_i(\mathbf{p}, \mathbf{T}, \alpha, \beta) = \frac{1}{1 + \exp(-\beta \mathbf{u_i R_\alpha Tp})}, \qquad i = 1, 2 \tag{5}$$

In equations 2 to 5, $\theta = (\mathbf{t}, \theta_{DS}, \theta_{ES})$, θ_{DS}, and θ_{ES} are the set of parameters of the TSE, the DS, and the ES functions, respectively, $\mathbf{T}$ is a translation matrix, $\mathbf{R}_\varphi$ is a rotation matrix of angle φ, $\mathbf{u_1} = (1, 0, 0)^T$, $\mathbf{u_2} = (0, 1, 0)^T$, e_x and e_y are the semiminor and semimajor axis of the central ellipse, β_e is the slope of the sigmoid curve that forms the central ellipse boundary, α_i is an angle with respect to axis i, β_i

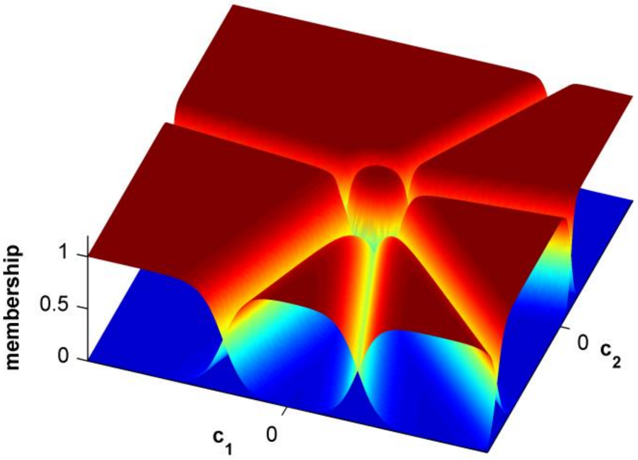


Fig. 5 TSE function fitted to the chromatic categories defined on a given lightness level. In this case, only six categories have memberships different than zero. (See color version of the figure at: http://www.cic.uab.cat/Publications/)

is the slope of a sigmoid function defined over axis i, and R_{α} is a rotation matrix of angle α.

Figure 5 shows an example of how the model divides a specific chromatic plane of the CIELab space.

3.2 Colour Constancy

Colour constancy is usually defined as the tendency of objects to appear the same colour even under changing illumination [28]. This is important due to the big variability of illumination in our real life (indoor/outdoor situations, midday/sunset daytime, etc.) For example, we will perceive as white a white piece of paper both in an indoor scenario or in an outdoor scenario at midday. However the information reaching the eye will be yellowish in the first case (tungsten illumination) and bluish in the second one. Several studies widely agree that human colour constancy is not based on a single mechanism [29].

In computational colour we simplify the human colour constancy property to convert it into a tractable problem. In particular, computational colour constancy tries to convert the captured scene under an unknown illumination into the same scene viewed under a white illumination (that is, we suppose that under white light, the perceived colours mimic the physical values). From a mathematical point of view, the problem is regarded as the search of a 3×3 matrix. However, for simplicity, researchers have widely used the Von Kries model [30] to simplify the problem. Von Kries model states that illumination change is a process which operates in each sensor response channel independently. Then, the 3×3 original matrix is converted to a diagonal one, greatly simplifying colour constancy computation. Mathematically, let us suppose we have an object with reflectance $S(\lambda)$ viewed under two illuminants $E_1(\lambda)$, $E_2(\lambda)$, and captured by a camera with sensitivities $R_i(\lambda)$, $i \in \{1,2,3\}$. Then, the colours captured by the camera are denoted as $\underline{\rho}^1$ and $\underline{\rho}^2$, where their components are given by

$$\begin{aligned} \rho_i^1 &= \quad S(\lambda)E_1(\lambda)R_i(\lambda)d\lambda \\ \rho_i^2 &= \quad S(\lambda)E_2(\lambda)R_i(\lambda)d\lambda \end{aligned} \tag{6}$$

Then, in computational colour constancy we search for α, β, and γ fulfilling

$$\rho^1 = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \gamma \end{pmatrix} \cdot \rho^2 \tag{7}$$

There are several methods trying to solve for this equation. The simpler ones (that actually give quite good results in real databases) are Grey-World [31] and MaxRGB [32]. Basically, GreyWorld assumes that the average of the scene is grey, while MaxRGB assumes the highest intensity values of the scene as a white point. These two methods were generalised by Shades-of-Gray [33] where the Minkowski norm was added and Grey-Edge [34] where image derivatives were also added. Some other methods deal with physical properties, such as mutual reflections [37], highlights and shading [36], and specular highlights [35]. Finally, another set of colour constancy methods are probabilistic such as Color-by-Correlation [38] and Illumination-by-Voting [39].

Recently, a new voting method [40] has been defined. This method follows the category hypothesis: Feasible illuminants can be weighted according to their ability to anchor the colours of an image to basic colour categories. In particular, it chooses the focals of colour names to behave as anchor categories. In this way, it returns as a solution the scene maximising the number of nameable colours. For example, if we have an outdoor scene in a field, it will return the image that converts both the sky and the green colours into the prototypical blue and green that have evolved with humans. Due to the naming nature of this approach, it would be the most suitable for our system, however, for limitations of the current mobile devices, a simpler method, the MaxRGB algorithm, has been used as a preprocessing step.

3.3 Properties of Sound

Physically, sound corresponds to mechanical vibrations transmitted through an elastic medium (gas, liquid, or solid) and is composed of longitudinal waves characterised by their frequency (or wavelength) and amplitude. Humans with normal hearing are capable of perceiving frequencies between 20 and 20,000Hz and intensities within a range of 12 orders of magnitude. When talking about sound, we refer to wavelength frequency as *pitch* and amplitude as *loudness* and interpret sound as a perceptual experience, in a way similar to how we interpret colour. When a key on a piano is struck, for example, we can identify both the pitch and loudness of the sound produced. The pitch is well defined and corresponds to physical properties of the wire struck (tension, linear mass density, and length), therefore we construct instruments manipulating these properties to produce different pitches. We can produce a louder sound by giving the key a bigger pull. In that case, the amplitude of the vibrations of the corresponding wire is bigger. Other attributes of sound events are *duration, spatial position* and *timbre*. Duration simply refers to the time span of a single sound event. On the other hand, the auditory system is capable of discerning the spatial localisation of a sound source. Localisation of sound events is by far less precise than localisation of objects by the visual system but not limited by the lighting conditions and in addition, hearing is omnidirectional.

By asking human subjects to tell the difference or express similarity judgement when listening to different sound excerpts corresponding to different musical instruments, one can derive timbre spaces. These spaces are perceptual and represent similarities between sounds. They are the counterpart in psychoacoustics of the perceptual colour spaces in vision, which are derived using psychophysics. However, giving a constructive definition of timbre is not easy and instead, timbre is often referred to a combination of qualities of sound that allow the distinction between sounds of the same pitch and loudness. To put it plainly, timbre is what allows us to tell the difference between a piano and a cello when both are playing the same note (pitch) with the same loudness (for the same duration and at the same position). Unlike pitch and loudness, which are characterised by frequency and amplitude, there is no single physical characteristic that directly relates to timbre. However, the main attributes of timbre are harmonic content and dynamic characteristics such as vibrato and the intensity *envelope* (attack, sustain, release, and decay).

3.4 Colour Sonification: Our Proposal

The central question is to find a systematic way to encode colour into sound. Such a mapping should have the following features:

i easy to use
ii not heady
iii coherent with synaesthesia (main features of)
iv perceptual isometry

Let us explain property (iv) in greater detail. Let C be a perceptual colour space and S a sound space. Suppose now that both spaces are endowed with a perceptual metric (denoted by $\ .\ _{\mathsf{C}}$ and $\ .\ _{\mathsf{S}}$, respectively). A mapping $\Phi : \mathsf{C} \to \mathsf{S}$ is said to be a *perceptual isometry* if the following property holds: for any two colours C_1, C_2 in C, if $\ C_1 - C_2\ _{\mathsf{C}} = T_{\mathsf{C}}(C_1, C_2)$, where $T_{\mathsf{C}}(C_1, C_2)$ is the discrimination threshold in the region of C_1, C_2 in C, then $\Phi(C_1) - \Phi(C_2) \;{}_{\mathsf{S}} = T_{\mathsf{S}}(\Phi(C_1), \Phi(C_2))$, where $T_{\mathsf{S}}(\Phi(C_1), \Phi(C_2))$ is the discrimination threshold at $\Phi(C_1)$ in S. Such a property would ensure no loss of discriminative power in the translation of colour into sound.

The first step in the construction of a timbre space is the extraction of physical characteristics. Sound events are expressed in terms of several time-frequency representations (harmonic sinusoidal components, short-term Fourier transform, energy envelope). Next, a large number of descriptors are derived which capture spectral, temporal, spectrotemporal, and energetic properties of sound events [43]. The information provided by these descriptors is highly redundant. Often, multidimensional scaling is applied to the space of descriptors to get a 3D space. The acoustic correlates of the three dimensions vary from a proposal to another. The spectral centroid receives a wide support in the literature and is often considered as the first and principal dimension (see [44] for a review on this issue). Another important dimension is provided by the attack time. The temporal variation of the spectrum is often adopted as the third dimension, but is less consensual. Note that describing sound using a three dimensional space S is a requisite if we are to define a perceptual isometry from a three dimensional colour space C to S. Both spaces should have the same dimension.

For computational reasons, we have implemented a simplified approach of the colour sonification which is mainly based on pitch for characterising sound. The input to the sonification algorithm is the output of the colour naming model described in section 3.1, that is, an 11-dimensional vector containing the membership values to the eleven colour categories considered. Hence, a colour is described by the 11 membership values of the colour naming descriptor.

In our approach, each chromatic colour category[1] is characterised by a different pitch (note) of a violin sound. The loudness of the sound is varied according to the membership value of the pixel to each colour category. To avoid noise, only membership values higher than 0.1 are considered. Therefore, given a colour, the generated sound will be a mixture of the sounds corresponding to the categories with membership values higher than 0.1, with different loudness each.

To differentiate between chromatic and achromatic[2] categories categories, timbre is used. Thus, achromatic colours are converted to a violoncello sound instead of the violin sound used to represent the chromatic categories. The differentiation among the three achromatic categories is done by assigning a specific pitch (note) to each of them: black is mapped to note C (do), grey is mapped to F (fa), and white is mapped to B (si). Table 1 summarizes the colour sonification scheme used.

Finally, the lightness of the colour, which depends on the value of CIELab coordinate L, is represented by different octaves. Hence, the lightness axis L is divided in two parts (low/high lightness) and colours in each part are represented by sounds on a specific octave.

[1] Red, green, yellow, blue, brown, purple, pink, and orange.

[2] Black, white, and grey.

Table 1 Summary of the conversion provided by the colour sonification algorithm

Colour	Pitch (note)	Timbre (instrument)
pink	E	violin
purple	D	violin
blue	C#	violin
green	A	violin
yellow	G#	violin
brown	G	violin
orange	F#	violin
red	F	violin
white	B	violoncello
grey	F	violoncello
black	C	violoncello

4 A Multimodal Device for the Visually Impaired

The mobile application developed is called **Coloresia** (i.e. a mixture between the words color and synaesthesia) and has three main modules, which are implemented as an Android activity[3]. WelcomeAct shows the initial interface of the application, Color2Sound is the main activity of the application and performs most of the tasks, such as acquiring images from the camera, displaying information on screen, or playing sounds, and ConfigAct allows the user to control the configuration of the application. Figure 6 shows a module diagram of the three activities of the application.

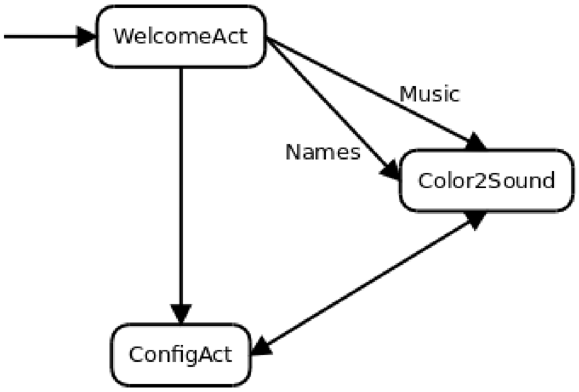


Fig. 6 Schematics of the main modules of the mobile application Coloresia

[3] In the Android platform, activities denote the basic components of applications. An activity corresponds to an interface of the application where the user can do some actions.

When the application is started, the user accesses to WelcomeAct, the initial activity of the application, which presents three buttons to the user. Two of these buttons take the user to the colour identification application in the two available modes, namely, music and voice. The third button calls the configuration module where the user can set different parameters of the application.

Figure 7(a) shows the interface of this initial activity. As it can be seen, the interface has been designed to facilitate the accessibility to visually impaired people: a large size font and colours with high differences in lightness have been used to highlight the text and make it easy to read.

From the WelcomeAct activity, the user can access to Color2Sound, the main activity of the application. When Color2Sound is started, the application acquires a sequence of images with the device camera and displays them on the screen. On one out of two frames of the sequence a region of interest (ROI) on the center of the image is selected. The dimensions of the ROI can be set by the user in the configuration activity.

The pixels' values in the ROI are averaged to obtain the mean RGB of the region. This mean RGB is the input to the colour naming method explained in section 3.1 to obtain the 11-dimensional vector with the membership values to the 11 colour categories considered. Then, this 11-dimensional vector is the input to the colour sonification algorithm presented in section 3.4.

Finally, the result of the conversion algorithm, i.e. a sound defined as a mixture of notes played by one or two instruments, is played on the device to allow the visually disabled users to know the colour of the objects at the center of the images they are acquiring with their device.

Besides the final sound played by the application, it also provides some information displayed on the screen of the device. This information is:

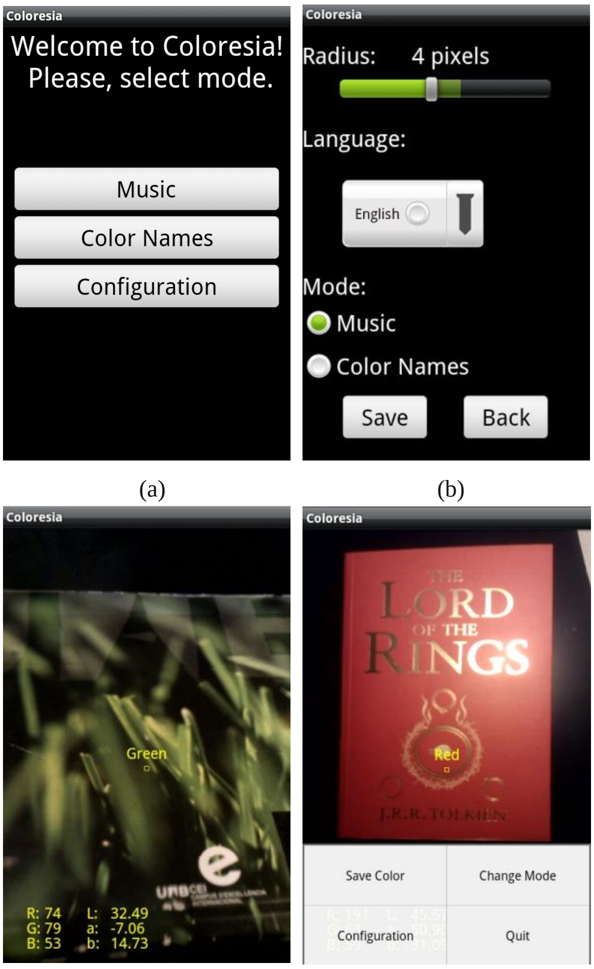
- The rectangle containing the region of interest.
- The colour name with the highest membership value corresponding to the mean RGB in the ROI.
- The mean RGB and CIELab values in the ROI.

Figure 7(c) shows the interface of the Color2Sound activity with all the information displayed on screen while the activity is working.

The Color2Sound activity also captures the events generated by the user on the touch screen. While this activity is working, the user can move the ROI through the image to identify the colour of a different image area. The user can also modify the size of the ROI, which can be set between a minimum size of 4×4 pixels and a maximum of 16×16. The size of the ROI can also be modified at the configuration activity as detailed below.

The user can also access the application menu from the menu key of the device. The options in this menu allow the user to save images on the device memory card, to access the configuration tool, to change the operation mode, and to exit the application. Figure 7(d) shows a screen shot of the menu layout.

The last module of the application is the configuration activity ConfigAct. In this activity, the user can set the three main parameters of the application. The first one is


Welcome to Coloresia!
Please, select mode.

Radius: 4 pixels

Language: English

Mode:
● Music
○ Color Names

R: 74 L: 32.49
G: 79 a: -7.06
B: 53 b: 14.73

(a) (b)

(c) (d)

Fig. 7 Coloresia interfaces. (a) WelcomeAct activity. (b) ConfigAct activity. (c) Main interface of the Color2Sound activity. (d) Auxiliary menu of the Color2Sound activity. (See color version of the figure at: http://www.cic.uab.cat/Publications/)

the radius of the region of interest, with a minimum of 2 pixels (i.e. a 4 × 4 window) and a maximum of 8 pixels (i.e. a 16 × 16 window). The value of this parameter can be adjusted by means of a sliding bar.

The second parameter is the language of the application. The selected language will be used in all the messages at the interface and by the voice synthesiser. The selection can be done by a *spinner* among the three supported languages: English, Spanish, and German. By default, English is initially selected. If the device does not have the language selected by the user installed on the device, the application proceeds to its installation. If, for any reason, the installation is not possible, the application warns the user by a message on the screen.

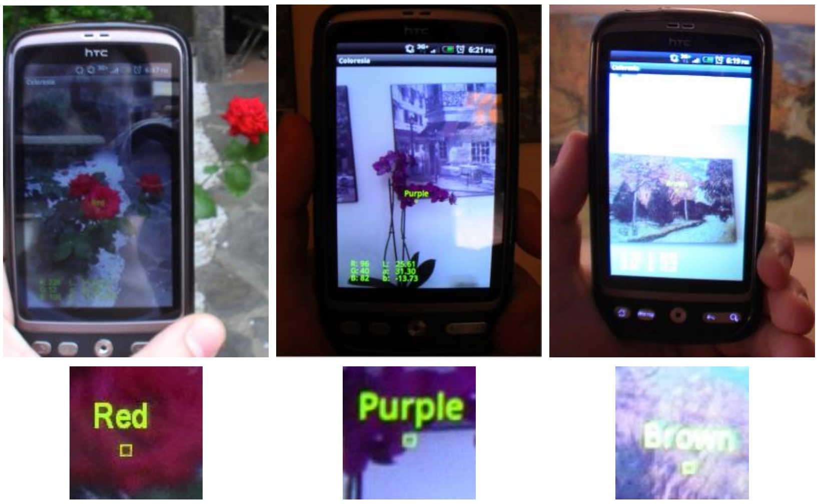
The third parameter that can be modified is the operation mode, where the user can choose between the default music output to represent the colours or a voice indicating the colour name of the stimulus detected by the application.

Finally, ConfigAct has two buttons to save the settings or going back discarding the changes. Figure 7(b) shows the layout of the activity that follows the same aesthetics as the previous activities.

4.1 Test and Results

The application has been tested on a HTC Desire mobile, with operative system Android v.2.2, a 1GHz processor, and 576Mb of RAM memory. The test of the application has been focussed on the processing time and the robustness against illumination conditions.

To test the speed of the colour identification part, the processing time of the 30 first detections on each test were averaged. The mean processing time was 123.18ms, with a standard deviation on 74.89ms. The test was only performed on the first executions to test the worst case, because after the initial colour detections the processing times reduce considerably to a mean processing time of 90ms.



(a) (b) (c)

Fig. 8 Examples of detections performed by the application. In the lower row, the central part of each image is zoomed. (a) Under natural daylight. (b) Under a reddish tungsten bulb light. (c) Under a mixture of daylight and tungsten bulb light. (See color version of the figure at: http://www.cic.uab.cat/Publications/)

Regarding the robustness against illumiantion changes, the application has been tested on three different illumination conditions: daylight, reddish tungsten bulb light, and a mixture of both. Although the application has more problems with low-illuminated environments, the application is able to correctly describe colours in most cases on the tested illumination conditions. Figure 8 shows three examples of the application working under the three illumination conditions.

5 Conclusions

In this chapter we have presented a prototype to help visually impaired people who is not able to see colour properly. The application is implemented on a mobile device and acquires images with the device camera. From this image, a region of interest is selected and the mean colour of the region is converted to a sound that is played by the device. Therefore, the users of this application are able to interpret these sounds and can identify the colours in the scene.

The method to represent colour as a musical sound is based on two steps. The first one transforms the input colour stimulus to a 11-dimensional vector representing the membership value of the colour to the eleven basic colour categories. The second step converts each membership value to a sound, and these sounds are combined to produce the final output of the system. From this output, the user can interpret the colour of the stimulus he or she has in front of.

With this application colour-blind people have an easy-to-use and low-price assistant for everyday tasks such as choosing the clothes to wear, understanding an underground map, or even interpreting a piece of art.

Acknowledgements. Authors are grateful for support from TIN 2010-21771-C02-1 and ConsoliderIngenio 2010 CSD2007- 00018 of Spanish MEC (Ministery of Science). They also acknowledge support from GRC 2009-669 of Generalitat de Catalunya. OP acknowledges Perfecto Herrera-Boyer and Emilia Gómez for their input during the preparation of the manuscript.

References

1. Foster, D.H.: Color constancy. Vision Research 51, 674–700 (2011)
2. Brown, R.O., MacLeod, D.I.A.: Color appearance depends on the variance of surround colors. Current Biology 7, 844–849 (1997)
3. Otazu, X., Parraga, C.A., Vanrell, M.: Towards a unified model for chromatic induction. Journal of Vision 10(12), article 5 (2010)
4. Berlin, B., Kay, P.: Basic color terms: their universality and evolution, Berkeley, Oxford (1969)
5. Genetics Home Reference: Color vision deficiency, National Library of Medicine, http://ghr.nlm.nih.gov/condition/color-vision-deficiency
6. Vision Problems in the U.S.: Prevalence of Adult Vision Impairment and Age-Related Eye Disease in America. Prevent Blindness America and the National Eye Institute (2008), http://www.preventblindness.org/vpus/2008_update/

7. Derefeldt, G., Swartling, T.: Color Concept Retrieval by Free Color Naming - Identification of up to 30 Colors without Training. Displays 16, 69–77 (1995)
8. Yendrikhovskij, S.N.: A Computational Model of Colour Categorization. Color Research and Application 26, S235–S238 (2001)
9. Seaborn, M., Hepplewhite, L., Stonham, J.: Fuzzy colour category map for the measurement of colour similarity and dissimilarity. Pattern Recognition 38, 165–177 (2005)
10. Mojsilovic, A.: A computational model for color naming and describing color composition of images. IEEE - Transactions on Image Processing 14, 690–699 (2005)
11. Benavente, R., Vanrell, M., Baldrich, R.: Parametric fuzzy sets for automatic color naming. Journal of the Optical Society of America A 25, 2582–2593 (2008)
12. Menegaz, G., Troter, A.L., Sequeira, J., Boi, J.M.: A discrete model for color naming. EURASIP J. Appl. Signal Process. 2007(1), 113 (2007)
13. Wang, Z., Luo, M.R., Kang, B., Choh, H., Kim, C.: An Algorithm for Categorising Colours into Universal Colour Names. In: 3rd European Conference on Colour in Graphics, Imaging, and Vision (2006)
14. Hansen, T., Walter, S., Gegenfurtner, K.R.: Effects of spatial and temporal context on color categories and color constancy. Journal of Vision 7 (2007)
15. Moroney, N.: Unconstrained web-based color naming experiment. In: SPIE Color Imaging VIII: Processing, Hardcopy, and Applications (2003)
16. Boynton, R.M., Olson, C.X.: Salience of Chromatic Basic Color Terms Confirmed by 3 Measures. Vision Research 30, 1311–1317 (1990)
17. Hardin, C.L., Maffi, L.: Color categories in thought and language. Cambridge University Press, Cambridge (1997)
18. Webster, M.A., Kay, P.: Individual and population differences in focal colors. In: MacLaury, R.E., Paramei, G.V., Dedrick, D. (eds.) Anthropology of Color: Interdisciplinary Multilevel Modeling, pp. 29–54. J. Benjamins Pub. Co., Amsterdam (2007)
19. Maurer, D., Pathman, T., Mondloch, C.J.: The shape of boubas: sound-shape correspondences in toddlers and adults. Developmental Science 9, 316–322 (2006)
20. Lewis, J.W., Beauchamp, M.S., DeYoe, E.A.: A comparison of visual and auditory motion processing in human cerebral cortex. Cerebral Cortex 10, 873–888 (2000)
21. Visual Music by Maura McDonnell (2002), `http://homepage.tinet.ie/˜musima/visualmusic/visualmusic.htm`
22. Neil Harbisson. Sonochromatic cyborg, `http://www.harbisson.com` (cited July 01, 2012)
23. Google Play, `http://play.google.com/store` (cited July 01, 2012)
24. Evans, B.: Foundations of a visual music. Computer Music Journal 29, 11–24 (2005)
25. Cronly-Dillon, J., Persaud, K., Gregory, R.P.F.: The perception of visual images encoded in musical form: a study in cross-modality infomration transfer. Proc. Roy. Soc. B 266, 2427–2433 (1999)
26. Bologna, G., Deville, B., Pun, T., Vickenbosch, M.: Transforming 3D coloured pixels into musical instrument notes for vision substitution applications. EURASIP J. Im. Video Process. 2007, 76204 (2007)
27. Rossi, J., Perales, F.J., Varona, J., Roca, M.: COL.diesis: transforming colour into melody and implementing the result in a colour sensor device. In: International Conference on Information Visualisation (2009)
28. Hurlbert, A.: Colour constancy. Current Biology 21(17), 906–907 (2007)
29. Hurlbert, A., Wolf, K.: Color contrast: a contributory mechanism to color constancy. Progress on Brain Research 144 (2004)

30. Worthey, J.A., Brill, M.H.: Heuristic analysis of von kries color constancy. Journal of the Optical Society of America A 3, 1708–1712 (1986)
31. Buchsbaum, G.: A spatial processor model for object colour perception. Journal Franklin Institute 310, 1–26 (1980)
32. Land, E.H.: The retinex. American Scientist 52, 247–264 (1964)
33. Finlayson, G.D., Trezzi, E.: Shades of gray and colour constancy. In: Color Imaging Conference (2004)
34. van de Weijer, J., Gevers, T., Gijsenij, A.: Edge-based color constancy. IEEE Transactions on Image Processing 16, 2207–2214 (2007)
35. Lee, H.: Method for computing the scene-illuminant chromaticity from specular highlights. Journal of the Optical Society of America A 3, 1694–1699 (1986)
36. Klinker, G., Shafer, S., Kanade, T.: A physical approach to color image understanding. International Journal of Computer Vision 4, 7–38 (1990)
37. Funt, B.V., Drew, M.S., Ho, J.: Color constancy from mutual reflection. International Journal of Computer Vision 6, 5–24 (1991)
38. Finlayson, G.D., Hordley, S.D., Hubel, P.M.: Color by correlation: A simple, unifying framework for color constancy. IEEE Transactions on Pattern Analysis and Machine Intelligence 23, 1209–1221 (2001)
39. Sapiro, G.: Color and illuminant voting. IEEE Transactions on Image Processing 21, 1210–1215 (1999)
40. Vazquez-Corral, J., Vanrell, M., Baldrich, R., Tous, F.: Color Constancy by Category Correlation. IEEE Transactions on Image Processing 21, 1997–2007 (2012)
41. Changeux, J.P.: Du vrai, du beau, du bien: Une nouvelle approche neuronale, Odile Jacob (2010)
42. Ward, J., Huckstep, B., Tsakanikos, E.: Sound-colour synaesthesia: to what extent does it use cross-modal mechanisms common to us all? Cortex 42, 264–280 (2006)
43. Peeters, G., Giordano, B.L., Susini, P., Misdariis, N., McAdams, S.: The Timbre Toolbox: extracting audio descriptors from musical signals. Journal of the Acoustic Society of America 130, 2902–2916 (2011)
44. Herrera-Boyer, P., Klapuri, A., Davy, M.: Automatic Classification of Pitched Musical Instrument Sounds. Signal Processing Methods for Music Transcription, Part II, 163–200 (2006)