Did I Really See That?

The Complex Relationship Between the Visual Stimulus and Visual Perception

John H. Krantz, Ph.D.

Hanover College

Paper in press; Do not quote without permission

Correspondence to:

John H. Krantz, Ph.D.
Hanover College
P.O. Box 890
Hanover, IN 47243
Email: krantzj@hanover.edu
Phone: 812-866-7316
Fax: 812-866-6752

Based on a talk of the same name presented at The Voice Foundation's 35th Annual Symposium:

Care of the Professional Voice on June 2, 2006 in Philadelphia, PA.

Acknowledgements

The author would like to express his deep gratitude both to Paul Reinert at Kay Pentax for his help in obtaining the laryngeal images used in this paper and to the reviewers of the drafts of this paper. Their help has greatly improved this paper. All errors remain mine alone.

Abstract

Laryngeal imaging uses optical and electronic means to visualize the larynx. Understanding some of the issues related to how the human visual system operates and how imaging systems interact with the visual system can help clarify some of the artifacts that arise from these technologies. This paper describes how the visual system can construct coherent perceptions from limited information, how it adjusts to current situations, and how the perception of any one part of the image depends upon the light levels around each point. In particular, the limited field of view and stroboscopic nature of the images can lead to many distortions from laryngeal imaging. This paper also describes the way that imaging systems sample the image, and the lack of stability inherent in an imaging system. The paper concludes with some observations and recommendations to improve the ability to use imaging systems in the diagnosis of laryngeal pathology.

Keywords:

Vision; Laryngeal Imaging; Imaging Systems

Did I Really See That?

The Complex Relationship Between the Visual Stimulus and Visual Perception

Laryngeal imaging has proved to be a powerful tool in the diagnosis and treatment of the voice[1]. The techniques for imaging the larynx have developed rapidly in the past few decades[2,3]. Despite all of the technological advances, diagnostic problems remain. Problems are associated with visual artifacts in the image[4], inter-observer agreement[1], and diagnostic specificity[5]. Many psychological factors contribute to these problems, including problems of decision making[6] and lack of training in visual thinking[7]. However, one important psychological factor is how an imaging system such as a stroboscope interfaces with the human visual system. A proper appreciation of how both the visual system and imaging systems operate can improve the critical assessment of laryngeal images. Such a critical viewing of laryngeal images can help the clinician avoid artifacts and improve diagnostic quality[8].

First and foremost, it is important to understand that the visual system exists to solve a problem and not as some idealized recorder of the world around us. The brain does not have any direct access to the world or even the rest of the body encased as it is in the skull. The sensory systems were evolved in parallel with the brain to meet the need of the brain to facilitate survival of the organism. What is implied in this statement is that sensory systems function to solve a particular problem for a particular brain. As a result, different animals will have different experiences of the visual world. For example, frogs have a rather primitive visual system with limits in sensitivity to details, color and even types of motion. Still, the frog visual system serves their needs. With all of our far more sophisticated visual and motor systems, catching a fly is a task that is very difficult given the human visual system, while a frog's visual system is uniquely designed to allow them to capture a fly out of the air easily with their tongue[9].

Imaging systems, such as a stroboscope, are also designed to solve a similar set of problems as our visual system. Both imaging systems (e.g., a camera) and visual systems (e.g., an eye) must capture an image and further process it (such as using a computer or the visual cortex). Yet, an imaging system is not self-contained. It must display that image usually on a monitor, so that it can be picked up and used by our visual system. Imaging systems must contend with the demands of the physical world to capture the image, the limitations of its processing capability, and the constraints placed upon those systems by having to interface with the human visual system. It is sometimes not possible to meet all of these demands, which is one source of artifacts in laryngeal imaging systems. To understand the issues raised by an imaging system-visual system interface, this paper will briefly examine how the visual system operates and then some generic issues about imaging systems' functioning.

Vision

The visual system captures electromagnetic radiation in a limited range of wavelengths. This radiation allows the brain to determine many complicated features about the world. While the brain mechanisms involved in the decoding of all of the information from light are important, they are beyond the scope of this paper. Instead, after a brief examination of the anatomy of the eye, most of this section will examine some of the ways our visual system operates to determine fundamental features of the world such as color and depth.

The scientific way to understand the visual system is carefully to manipulate the stimulus and determine how perception changes as a result. In describing what has been learned about perception as a result of these studies a similarly interactive approach is preferred[10]. To realize this goal, interactive versions of many of the figures found herein are also found at the website developed from the presentation delivered at the 35th Annual Meeting of The Voice Foundation upon which this paper is based.[1]

The eye

As is well known, light enters the eye, is focused by the cornea, lens and pupil, and is imaged on the retina. The retina is a paper-thin layer of cells at the back of the eye where the process of seeing begins. The retina is made up of several layers of interconnected cells. The physiological act of seeing begins in the layer of the retina farthest from the front of the eye where the receptor cells are found. The retina has two types of receptor cells, rods and cones that absorb the photons of light. The rods are more sensitive to light and work under conditions of very low light levels. The cones are less sensitive and function in daylight. In addition, there are three classes of cones which is the basis of color vision. It is the daytime vision, based upon the

[1] The address of the website is: http://psych.hanover.edu/presentations/VoiceFound/

functioning of cones, that is relevant in laryngeal imaging. The nature of vision that is associated with the functioning of rods will not be discussed.

The retina is not homogenous over the back of the eye. Two extreme points of variation in retinal structure and functioning are the optic disk and the fovea. The optic disc is where the optic nerve leaves the eye and a small part of the blood supply for the eye enters. As a result, there is no vision in this part of the retina which is why it is also known as the blind spot. It is located about 20 deg. towards the nose from the fovea. The fovea, in many ways, is the center of the retina. When you say you are “looking at” some object, the image of that object is falling on the fovea. The fovea is the region of the retina with the greatest acuity during the daytime. The fovea does not cover much of the retina, much less than 1%. But nearly 50% of our visual cortex that processes the visual information from the eye responds to information from the fovea[11]. The foveal acuity has a great deal to do with the density of the cones and other cells processing information from these cones. The cones are packed far more densely in the fovea than at any other part of the retina[12]. The relevance of the small size of the fovea to laryngeal imaging is that it is impossible to examine an entire image in detail at one time. Figure 1 gives one example of a laryngeal image to view for purpose of the present discussion. This or any other image must be scanned with the eye moving over the image and the image reconstructed from successive fixations. A great deal of work remains to be done to understand all of the details about how this reconstruction is done, but its importance is not to be underestimated.

Perception

Perception is often defined as the meaningful interpretation of the basic sensory data. This definition contrasts perception with the basic reception of sensory signals. Most of the phenomena that can lead to misinterpretations of a laryngeal image arise in perception as currently defined, so the focus of the following discussion will be on these activities of our visual system.

Themes. There are three themes about visual perception that are important to keep in mind. First, the visual system actively organizes the sensory information. It is not a passive recipient of the images from the world but is influenced by past experience, meaning, and expectations[13]. Second, perception is contextually based. How any one point in the world appears depends not just on the light from that location but also on the light coming from around that location. Third, the visual system is dynamic in function. Under different conditions, the visual system organizes itself differently to optimize its functioning. The differences between the use of cones in the daytime and rods at night is just one example. As a result, the way some object is perceived may well depend upon the environmental conditions and its impact on visual system organization. These themes will be weaved in and through a discussion of vision along several basic dimensions of perception.

Spatial. Examine Figure 2 and try to determine what is depicted. Most people report seeing a triangle in the middle of the lines and circles. The figure really has lines and circles with triangular slices removed. There is no triangle. Yet, most people report clearly seeing a triangle. This apparent triangle has several interesting properties. It is seen in front of the background of which it is a part. It is seen as brighter than the surrounding white even though there is only one white on the figure. Clear contours seem to form the three edges. This illusion is called the Kanizsa triangle and is the classic demonstration of illusory contours[14]. In this example, the active way that the visual system organizes input is brilliantly shown. From a few isolated stimulus features, a triangle is perceived.

The Kanizsa triangle allows us to postulate several rules which the visual system uses to organized stimuli. First, the visual system tends to organize perception into a figure in the foreground and the rest of the scene tends to play background. This feature of perception is known as figure-ground perception. In the Kanizsa triangle, the need for a figure is so strong, that it is formed out of a minimal number of features, it could be just the three circles in the corner, and the rest is filled in. Second, the visual system seeks the simplest possible interpretation of the scene. In addition, the visual system prefers the perception of physically stable arrangements, when possible. Seeing a triangle is simpler and more stable in a physical sense than seeing several disconnected lines and three odd shapes.

This tendency to organize visual input into the simplest and most stable form possible can both assist and impair diagnostic viewing of laryngeal images. The ability to organized disparate elements can facilitate seeing features of the motion of the laryngeal wave that are distributed across the vocal folds as being related together. For example the mucosal wave that rides on both sides of the vocal folds is perceived as moving as a single wave partly through this organizational tendency of the visual system. Conversely, and particularly in degraded images, these same organizational principles of the visual system could suggest pathologies that do not exists much as the triangle that does not exist is perceived.

The spatial layout of an image allows the contextual nature of the visual system to be revealed. Figure 3a presents the phenomena known as simultaneous contrast. The two central squares are both exactly the same. However, the square on the left, surrounded by black, looks brighter than the square on the right, surrounded by white. It turns out that how bright any part of a scene appears depends as much on the light intensity of the areas around the region as the intensity of that region itself[15]. There is a related color illusion where surrounding colors alter the appearance of a central color. How any location in any scene appears is greatly altered by the surrounding context.

Applying this finding to laryngeal imaging, it is important to be careful when judging the brightness of any region of the larynx, say an apparently inflamed region, as its appearance depends not only on the brightness of that region but the surrounding regions. Figure 3b shows a laryngeal image that illustrates this point. The light area indicated by the 1 is actually the same luminance values and even slightly less intense as the dark area indicated by 2. However, the area around the patch indicated by the 2 is much more intense than the area around the patch indicated by the 1. These surrounding areas change the brightness of the two regions. An area of investigation would be whether the absolute levels of intensity are more diagnostic than the brightness of the different areas as seen by the eye. In the former case, using the software processing that is available on any computer would be an important diagnostic tool. In the latter case, this same software could be often misleading and visual inspection would be more important at least until tools became available that processed images more like eyes do.

One emerging area of research is the role of attention in perception, particularly in the recognition of forms. Some aspects of perception, such as the organizational tendency mentioned above[16], are not influence by how the observer attends to the image. However, other features of the stimulus, such as the beginning of motion, can grab attention from other, perhaps more important, features of the stimulus environment[17]. The role of attention in other features of form perception are areas of active debate[18,19]. Much is yet to be learned about how the way an observer uses attentional resources can alter what is perceived. The resolution of such debates could have great implications for training users of laryngeal imaging.

Chromatic. The ability to see in color adds a great richness to our experience of the visual world, e.g., the rich array of colors of a sunset on an ocean. Diagnostically, many disorders of the larynx are accompanied by irritation which shows up most easily as a change in color. The presence of color vision also adds a great deal of complexity to the visual system. Color reproduction technology such as film, photographs and color printers are based on how our color vision operates. Since most laryngeal imaging systems use color reproduction based on the human visual system, it is important to understand this aspect of the visual system.

A fundamental principle of color reproduction technologies is that they rely upon the ability to mix two or more colors to get a different color. So a small set of primary colors, usually three, is chosen to create an extremely wide range of other colors. There are two basic ways that color mixing is done. Subtractive color mixing usually is done when the pigments being used to mix colors do not generate their own light. The pigments are illuminated by an external light source, ideally a white light source composed of a very broad array of wavelengths. When the light hits the pigment, some of the wavelengths are reflected and others are absorbed. For example, a red pigment absorbs most wavelengths that hit it except for long wavelengths and a blue pigment absorbs most wavelengths except short wavelengths. As indicated by the ink cartridges of a color printer, the three primary colors that best serve subtractive color mixing systems are cyan, magenta, and yellow. Additive color mixing usually occurs when the device can generate its own light and not merely reflect light from an external light source, e.g., a color monitor. The primary colors are placed so close together on a screen that they cannot be resolved by the eye but are blended together by the visual system. Examining the surface of a monitor with a jeweler's loupe or strong magnifying glass will reveal the primary colors that are used to generate all of the colors on the screen. If three primary colors are used, as is typical, the primary colors will be red, green, and blue.

The study of color mixing goes far back into history. The research and informal work on color mixing has led to the development of numerous color wheels to summarize the regularities observed. In many ways, the work on color mixing culminates in the work of the Commission Internationale de L'Eclairage (CIE) color mixing system first published in 1931[20]. Figure 4 shows one representation of the CIE system. This representation shows the range of colors possible to be seen in what are called chromaticity coordinates. In essence, this figure is the CIE's color wheel. To develop their system, it was necessary to define the standard color observer. The CIE system allowed for automated color mixing which makes possible color film, color television, color monitors and color printing. This system represents an enormous technical achievement but it works only for the standard color observer, which does represent most but not all of us.

While the details of the physiology of how the visual system processes color were unknown at the time, the CIE color matching process derives its power from how the visual system handles color. Figure 5 shows the first stages of color processing in the eye. Humans have three classes of cone receptors. Each class of cone is most sensitive to a different range of wavelengths of light. These cones are often called the blue, green and red cones but a better and more accepted set of names are short wavelength, middle wavelength and long wavelength cones, respectively. The number of primary colors that is minimally necessary to match a wide range of colors is determined by the number of cone classes, which is why this theory of color vision is called the Trichromatic Theory.

As shown in Figure 5, the cones are the only part of the visual system that are trichromatic. After the cones, the color information is processed according to the color opponent theory. The cells in the visual system that process color in a color opponent fashion exist at several levels of the visual system. So the connections from the cones to the color opponent cells indicated in Figure 5 are conceptual in nature and do not represent explicit connections or cells in the visual system. The color opponent processing in the visual system plays a major role in giving colors much of the quality of color experience. While it takes three primary colors to make a match for a wide range of other colors, colors are experienced as one or a mixture of two of four unique hues. These unique hues are red, green, blue and yellow. It is possible to actually, with a broad brush, describe all colors with these four hue names[21]. In the color opponent cells, these unique hues are arranged in oppositional pairs, i.e., mixing one of a pair with its opposite tends to cancel the perception of color leaving a neutral color of white, gray or black. There are two sets of oppositional pairs, Red and Green (R/G) and Blue and Yellow (B/Y). When a red wavelength strikes the eye, it is mostly absorbed by the long wavelength cone. This excites the R/G cell which signals red back to the brain. When a green wavelength strikes the eye, it primarily excites the middle wavelength cone which in turn inhibits the R/G cell, sending a green signal back to the brain. A blue wavelength is principally absorbed by the short wavelength cone.

The short wavelength cone, in turn, stimulates the B/Y cell which sends a blue signal back to the brain. A yellow wavelength stimulates both the long and the middle wavelength cones about equally. These cones cancel each other out on the R/G color opponent cell so that neither a red nor a green signal is sent to the brain. However, both the long and the middle wavelength cones inhibit the B/Y color opponent cell which sends a yellow signal to the brain. Thus, three primary colors in the trichromatic system turn into four unique hues in the color opponent system. For completeness, it is important to mention that there are also green excitatory/red inhibitory R/G color opponent cells and yellow excitatory/blue inhibitory color opponent cells. They act in exactly the opposite way described.

The description of human color vision, thus far, has been for the standard normal observer specified in the CIE system. However, within the population, many people have color vision that varies from the standard normal observer. Some of the variations are subtle, others profound. Many of these variations from the color normal are often miscalled color blindness. True color blindness does exist but it is quite rare. Dichromacy will be used to illustrate one variation from the color normal. In dichromacy, one of the three cone classes is missing. If the long wavelength cone is missing, the person is said to have protanopic vision, if the middle wavelength cone is missing, the person is said to have deuteranopic vision, and if the short wavelength cone is missing, the person is said to have tritanopic vision. People with protanopic or deuteranopic vision are both said to have a form of red-green color blindness. Due to the way that both long and middle wavelength cones contact the R/G color opponent cells, these people tend to confuse reds and greens. People with tritanopic vision tend to confuse blues and yellows, as can be seen from the way that the short wavelength cone connects to the B/Y color opponent cells and are said to have blue-yellow color blindness. Deuteranopic vision is the most common form with tritanopic vision being by far the rarest form of dichromacy.

Figure 6 shows a montage of a healthy laryngeal image, first normally, and then modified to simulate all three forms of dichromacy[22]. Notice, first, the almost complete loss of color for the person with protanopic vision (Figure 6b). This loss of color results from the fact that laryngeal images are composed largely of long wavelengths (reds). By losing sensitivity to these wavelengths, most of what gives color to the image is not being detected by someone with protanopic vision. The situation is similar but not identical for a person with deuteranopic vision (Figure 6c). Since both long and middle wavelength cones input in to the R/G color opponent system (Figure 5), loss of either one of the cones tend to both cause loss of color appearance in related ways. A person with tritanopic vision see color very differently (Figure 6d). Again, since laryngeal images are mostly composed of long wavelengths, the loss of the short wavelength cones causes very little change in the image for the person with tritanopic vision. There is some loss of the blues that are in the normal image, but that has less of an impact, overall, than what results from the loss of the other two cones.

These disorders are very important in laryngeal imaging because these differences in the processing of color impact the way colors appear on all monitors. A person with dichromatic vision will not be able to nearly as effectively able to use color in diagnosis when viewing a monitor as will a person with trichromatic vision because the color reproduction is designed for the color normal observer, the person with trichromatic vision. In direct examinations of the larynx using a mirror, variations in the observers color visual system will have less impact on the ability to use color in diagnosis. As a recommendation, it seems important to scan for color deficiencies and then direct and train observers who have color deficiencies to rely more on direct observation of the larynx when color is important for diagnosis.

Temporal. One of the most important reasons for the development of the different diagnostic tools for imaging the larynx is to better see the motion of the larynx. In particular, these techniques, via different methods, try to slow down the vocal fold vibrations so that our visual system can more easily process their movement. Whether the imaging technique used to evaluate vocal fold vibration is stroboscopy, kymography, or high-speed imaging, in all cases a set of still images is displayed on a monitor at a rate high enough to generate the appearance of real motion to the human visual system. Thus, all of the imaging methods rely upon the visual system’s ability to perceive still images as moving which is called apparent motion. Any illusions that can affect one imaging method can affect any other imaging method, in principle, although the rate at which the images are captured is important. The way that the visual system processes temporal information, then, is very important.

A limited analogy is possible between the visual system and how a camera and film work together to create an image. To have enough light to see, the visual system and film sums the light that falls on the retina over a period of time. In a camera, this period of time is defined by the shutter speed, which is the time during which the frame of film is exposed to light. In vision, where there is no shutter but continuous exposure to light, the eye still gathers light over a brief period of time. This collecting of light is referred to as temporal summation. The period of time during which temporal summation occurs is defined by Bloch's Law (defined as: Response = Intensity x Time). If an image is moving across the retina or the eye is moving faster than is defined by Bloch's Law, the image will appear blurred. Just as the shutter speed can be adjusted to optimize camera performance under different conditions, the visual system adjusts the duration of temporal summation for much the same reasons. The slower temporal summation of our night vision is responsible for the streaking of sparklers as they are waved through the night on the Fourth of July. Any speed of oscillatory movement of the vocal folds is much faster than any temporal summation used by the visual system. Thus, it is necessary to use very brief flashes to image the larynx to see all of the complex motions clearly.

Related to temporal summation is the Critical Fusion Frequency (CFF). The CFF refers to the rate at which a light source needs to flicker to be perceived as a continuous light. Most artificial light sources flicker. Televisions in the United States and most monitors flicker at 60 Hz. Movies will usually flicker at 48 or 72 Hz. A standard value for the CFF is given as 60 Hz and is used as a standard value in many engineering applications. However, the actual CFF for any environment actually depends upon many factors including the light level and the size of the image but the 60 Hz value works well for normal room light levels and images of the size of monitors[23].

It is the flickering nature of images shown on movies and monitors that leads to one of the classic visual illusions, the wagon-wheel illusion. This illusion gets its name from the way that wagon wheels in westerns can sometimes appear to rotate backwards on a screen while moving forwards. Since all forms of laryngeal imaging use flickering images, it is worthwhile to examine this illusion in some detail. Figure 7 shows an illustration of how the wagon-wheel illusion works. There are two points of time illustrated in Figure 7. At Time 1, the wheel is in the position indicated by the solid lines and at Time 2, the next time the image is shown, the wheel has rotated clockwise to the position indicated by the dashed lines. The solid arrows indicate the actual movement of the wheel. Since the image is stroboscopically displayed, there is ambiguity in the possible interpretations of the movement of the wheel. Since there is no visible motion between Time 1 and Time 2, it is possible that the wheel could be seen moving in the correct clockwise direction or the incorrect counterclockwise direction. From what was learned about the rules the perceptual system follows from the discussion of the Kanizsa Triangle, a reasonable hypothesis can be generated. The visual system tends to choose the simplest and most stable way to perceive a scene. In this case, it is simplest to see the scene where the spokes move the shortest distance. So, to look more closely at Figure 7, the motion of the wheel between Time 1 and Time 2 takes it to where the spokes move so far that they are closer to the next spokes position at Time 1 than where each spoke started. In this case, the simplest perception is of the incorrect counterclockwise motion, which is what would be perceived in this case.

There is one feature of Figure 7 that has been ignored so far. Attached to the outside of the wheel at the end of one spoke is a circle. It is drawn in black at Time 1 and in gray at Time 2. Using the same logic used to determine the perceived motion of the spokes, there is only one reasonable way for the visual system to perceive the motion of this circle which is in the correct clockwise motion. As a result, the dot will be perceived to move in the correct direction while the spokes are perceived to move in the incorrect direction. Note that when motion is generated from stroboscopic imaging, it is possible to perceive both the correct and incorrect direction of motion from the same scene. Of course, a faster strobe will allow fewer of these sorts of motion illusions to occur. Thus, high-speed imaging will have fewer motions appearing to move in the wrong direction than stroboscopy. Still, it is important to recall that real motion is not being displayed to the visual system and illusions are possible. As high-speed imaging advances, it might be possible to develop new diagnostic criteria from the relative motions of the different features of larynx motion such as the timing of the mucosal wave relative to various features of vocal fold vibration.

So far, the discussion has implied that the rate of flickering that is necessary to reach the CFF and to generate the perception of motion from a sequence of still images (apparent motion) is identical. However, that is not the case at all. Movies present new images at 24 Hz, flashing each image two or three times to reach the 48 Hz or 72 Hz flicker, respectively. In television, images are updated only at 30 Hz (U.S.) or 25 Hz (Europe). On the web, the update rate for video can vary widely but usually from about 10 Hz to 30 Hz. From these numbers it is clear that the flicker rate that is necessary to generate apparent motion is much slower than what is necessary to eliminate the perception of flicker. As the sharing of videos of laryngeal images increases and the variations of frame rate across media, it is important to establish standards for appropriate frame rates to accurately reproduce the important motions for diagnosis.

Depth. In laryngeal imaging, the examiner often tries to determine depth of features in the scene, including the relative depth of the vocal folds. These determinations of depth can play an important role in diagnostics[1]. As such, it is important to understand how the impression of depth is created from viewing a two-dimensional scene. In some ways, the visual system has the same problem interpreting depth off of a flat monitor surface as viewing the real world. The world is projected onto the two-dimensional retinas. The visual system tries to take advantage of as much of the information in a scene that indicated the absolute and relative depth of an object as possible. The sources of information about depth in a scene are traditionally called depth cues.

Some depth cues are present in the two-dimensional scene and can be used by one eye alone, and are called monocular depth cues. All laryngeal imaging that use a monitor is limited to monocular depth cues. A subset of monocular depth cues can be recreated on a still photograph or a painting, called pictorial depth cues. Figure 8 illustrates some pictorial depth cues. Instead of an exhaustive listing, only a few depth cues will be discussed. Almost any textbook on sensation and perception would give a good basic introduction into depth cues (eg, see Goldstein[24] for a more detailed discussion).

Looking at Figure 8, notice how the two lighter circles, the ones that appear farther away, are smaller than the two darker circles which appear closer. As an object gets farther away from the observer, the size of the retinal image of that object gets smaller; a depth cue called relative size. The change of the size of the retinal image as the object gets farther away is the result of the geometry of the situation, so it is a property of the distance between the object and observer. The brain is able to interpret this difference in retinal image size as a difference in distance.

The ability to use relative size as a depth cue illustrates a different type of context that impacts how we perceive objects. Sometimes to be able to use relative size as a depth cue, it is necessary to have some idea of the actual size of the object. This need for memory would be necessary if only one example of the object were present in the scene, e.g., only one circle in

Figure 8. Thus, past experience can play a role in influencing how we see objects. Given that visual experience is an important context for interpreting a visual scene, it is important that clinicians regularly see a wide range of laryngeal images, both healthy and not[3,4].

For the next depth cue, notice the objects are closer to the dividing line between the top darker region and the bottom lighter region. The eye easily interprets this dividing line as the horizon. In the depth cue known as relative height, the closer an object is in height to the horizon the farther away it appears. For the two objects below the horizon on the left side of the figure, the object that is higher appears farther away. This perception corresponds to how the ground of the scene rises from the feet of the observer to the horizon. For the two objects above the horizon, the object that is lower on the figure appears farther away. This part of relative height corresponds to how the sky descends to the horizon in any viewed scene. This depth cue is often misdefined only in terms of objects below the horizon. In most laryngeal images, not having the endoscope directly positioned in front of the larynx will create illusions of depth with the horizon being above or below the screen[4].

There are two interrelated depth cues illustrated by the pattern of lines forming the background. Consider just the floor of the scene. The lines that appear to be receding into the distance are perceived to be parallel. In actuality these lines are vertical on the page and are tilted towards each other. The same geometry that leads objects to have a smaller retinal image size the farther they are from the observer is involved here. This tendency for parallel lines to converge as they get farther away is known as linear perspective. The squares that make up the floor, walls, and ceiling of the background also reveal some interesting patterns that assist depth perception. Along the floor, side walls, and ceiling, the squares appear to get smaller in the distance. Along the back wall, the squares all stay the same size. The relationship of what happens to the size of the squares and changes in depth is called texture gradient. When the texture elements of the surface stay the same size from the perspective of the observer, the surface appears vertical and to be facing the observer like the back wall. When the elements change size, like the floor, the surface seems to recede in the direction that the texture elements get smaller. For laryngeal imaging, linear perspective and texture gradient are particularly important in regard to some of the distortions related to not having an endoscope properly positioned relative to the vocal folds. If the endoscope is not in proper alignment with the vocal folds, the parallel lines and barrel distortion problems will emerge, creating apparent changes in depth and shape that are not part of the actual scene[4].

One monocular depth cue that cannot be represented in a still picture is motion parallax. Motion parallax is a depth cue arises from our own motion. Imagine driving down the highway. Looking sideways out of the window, notice how the grass at the side of the road seems to pass by much faster than the mountains or trees in the distance. Just as the images of objects that are close are much larger on our retina than objects in the distance, so as we move, close objects cover a much larger distance across the retina than do distant objects. This relative motion information is picked up by the brain and interpreted as differences in depth. Motion parallax is particularly powerful in creating the impression of depth and can disambiguate information from other depth cues which can be difficult to interpret or even conflicting. However, there is little possibility of motion parallax in laryngeal imaging.

While monocular depth cues can create a very vivid impression of depth, binocular depth cues, which require two eyes, are particularly powerful but only over restricted ranges of depths. One binocular depth cue arises from the eye movement known as vergence. Recall that the fovea has the highest acuity on the retina. As a result, it is of vital importance that the two eyes be aligned so that when looking at an object, the same part of the object falls on both foveas. Even a small difference in the position of the images falling on the two foveas can seriously disrupt vision. Since, the eyes are not in the same position in the head, they must rotate relative to each other to properly align both foveas towards objects at different distances from the observer. Figure 9 illustrates this situation. The two eyes are examining the dot (**A**) connected by the dashed lines to the eye. To have the image of that point on both foveas, the eyes must be rotated so that the line connecting the point at **A** to each eye crosses the retina at the fovea. To then look at the object at **B**, the eyes must rotate inwards so that the lines from **B** will cross each retina at their respective foveas. The closer the object, the more the eyes rotate in, i.e., the more they cross. The sensory feedback from the eye muscles is the source of the vergence depth cue. For objects more than about approximately 20 feet away from the observer, the eyes are essentially parallel, determining an upper limit to the distance where vergence information can be used as a depth cue.

The rest of Figure 9 illustrates the binocular depth cue termed binocular disparity. The eyes are looking at the dot at **A**. By definition, this object falls on exactly the same positions on the retinas of both eyes, i.e., on the fovea. Only objects that fall on the arc that goes through this point also fall on corresponding locations of both retinas. This arc is called the horopter. Objects that are in front of or behind the horopter do not fall on corresponding points in the two eyes. Examine the object at **B**. Following the solid line from **B** to the left eye, the image for **B** falls to the left of the fovea (indicated by where the two lines cross the left retina). Following the solid line from **B** into the right eye, the image for **B** falls to the right of this fovea. If you look at the right eye, the position of the image from where **B** falls in the left eye is shown with the darker solid line. The image from **B** falls on very different locations on the retinas of the two eyes. This difference in the position of the images of **A** and **B** in the two eyes is binocular disparity. It is this depth cue that is being exploited in 3-D movies. Binocular disparity allows a very precise determination of depth and discrimination of depth differences, the finest possible. However, there is only a fairly limited range of disparity that can be used for a precise perception of depth. If disparity exceeds this limit, about 22 arcsec[2] on both sides of the horopter, as in Figure 9, only relative depth can be extracted and instead of a single image, a double image may be perceived[25].

---

[2] Arcsec are units of angle. Each degree of angle can be divided into 60 minutes of arc or arcmin. Each arcmin can be divided into 60 arcseconds. For reference, the full moon is about 30 arcminutes.

So binocular disparity is a depth cue that operates over a limited range of depths and it allows a very fine discrimination of depth in that range, such as seeing a step or detailed changes of depth in an object being manipulated. When viewing the larynx directly with a mirror, binocular disparity operates as a depth cue and may allow for finer depth discriminations than when viewing via an endoscope and monitor.

While depth discrimination is possible when imaging the larynx, the limited viewing angle possible can lead to illusions in the perception of size. The two mugs in Figure 10 are exactly the same size. The mug on the right is farther away from the camera, which makes it smaller on the picture, the geometry of relative size. The picture is taken from a very careful perspective that limits the richness of the visual information. First, the image is cropped to limit the depth cues, such as texture gradient, which make it harder to perceive that the mug on the right is farther away. In addition, the picture was taken at an angle that limits the relative-size depth cue. In movies, this limitation of depth cue and close attention to angle used to change apparent size is called forced perspective, and is one of the ways that the *Lord of the Rings* movies created the impression that the actors playing the hobbits are as short as hobbits are supposed to be. The easiest way to eliminate this illusion is to change the angle of view. The limited field of view in the larynx and the inability to move the endoscope both increase the likelihood of these types of illusions and limit the best way to overcome them.

Refer back to Figure 3b which shows a laryngeal image that will serve as a reference to discuss the issues of depth perception directly on laryngeal imaging. In this typical image, the lighting is directly on the larynx reducing shadows to a minimum. Interposition is most evident where the vocal folds hide what is under them for the most part except when the vocal folds are open, and the lack of field of view is evident. Shadow does play some role to highlight some laryngeal features such as the relative nearness of the epiglottis. On this surface, it is also possible to see some changes in texture gradient from the blood vessels which help give some depth information. In some clinical situations, it is important to be able to determine if the vocal folds are at the same depth[1]. If the folds overlap during approximation, it is easy to make this determination via interposition. If the folds just meet, however, without the capabilities of binocular vision, this difference in depth could be missed. Thus, it seems reasonable that for diagnoses of disorders where depth perception is relevant, direct observation methods that allow for binocular vision should be recommended.

Imaging Systems

Having this brief description of the visual system as a background, we can now turn to some of the generic issues that impact the accuracy of imaging systems. In many cases, the operation of the visual system will interact with the operation of imaging systems which can create artifacts. These imaging systems issues will be grouped around two main topics, the way that imaging systems sample information and the stability of the monitor.

Sampling

No imaging system, including our own visual system, can capture everything in the world that is revealed by light. So these imaging/visual systems sample the visual information along several dimensions. The goal is to capture and even emphasize needed visual information while minimizing the impact upon perception of the sampling. The following sections examine luminance, spatial and chromatic sampling.

Luminance. The visual system is capable of operating over an enormous range of light levels. If 1 is the lowest light level that can be perceived, then the brightest will be 10^{13} times more intense. This entire range of luminance cannot be processed at any one moment by the visual system. Still, the visual system can instantaneously operate in a light range well exceeding a luminance range of 100 to 1. To extend this instantaneous range, the visual system adjusts its sensitivity to match the prevailing light level. It is even possible for different regions of the retina to adjust to different light levels at one time[24]. On the other hand, imaging systems, in particular monitors, have a much more limited ability to reproduce differences in light levels. First, most color monitors can generate light levels only going from ~30 to 1.[20] Even this light level range only applies to differences between white and black on the screen. All other colors have a more limited range of light levels and cannot generate as bright a light level because of the additive way that colors are created on monitor screens. The blue primary is by far the least intense of the three[20].

Obviously, imaging systems cannot operate over nearly as great a range of luminance levels as the visual system. At a basic level, the implications of this limit are two-fold: 1) the luminance level of the light source must be matched to the ability of the image-capturing device and the monitor to handle and generate luminance, and 2) even when matched, it is still possible that the light range of the scene will exceed the monitor's ability to reproduce light levels, which may hide features of the scene. If the first condition is not met, the image will either be too dark or too bright, reducing the contrast in the image necessary to read it clearly. If the second condition exists, then portions of the image may be washed out while other parts of the image will appear clear. Figure 1, although an excellent image, still shows a few minor areas that illustrate this point. The bright area on the cuneiform tubercle in the upper left corner of the image is essentially a region at the maximal intensity for the monitor. Any distinctions in intensity in this region, small though it is, is lost and cannot be recovered. Conversely, the dark region behind the areyepiglottic fold (outlining the entrance to the pyriform sinus) on the right side of the image is all in the region of luminance values where monitors do not produce any actual luminance differences in most cases. Again, this information is lost; fortunately, it is not important in this case. In laryngeal imaging, it is possible to have a mismatch between light level from the light source and that needed by the endoscope and the monitor. This issue is particularly acute given the slow rate of dimming of the brightness of the endoscopic light source[4]. Carefully following the manufacture's calibration procedures are vital to getting well balanced luminance in the images. Despite this, however, it is probably not possible to capture all of the luminance information of an image. Thus, these calibration procedures should focus on those regions of the larynx that are of importance to the diagnosis under consideration. To rush this step in imaging is to render images that may not be useful for diagnosis.

The way that monitors sample the intensity of light within the range they produce is also important to understand. Since monitors are computer controlled, computers do not generate every luminance value of which the monitor is capable. Instead, only discrete luminance values are used. Most monitors have 8 bits (256) of luminance values for each primary. Some monitors now have 12 bits of luminance values. However, the light level produced by the monitor is not linearly related to the bit value used to generate the light level, but is better described by an exponential function known as the gamma. For example, the light level of 200 on the green primary is not twice as intense as the light level at 100, as would be expected in a linear system[8]. One important implication is in using an image-editing program to analyze the image. Such programs can collect all of the values used to generate the image, usually separated for each primary. However, these values do not linearly relate to the intensity of the image surface and should not be used in clinical judgments about characteristics of the larynx under examination without careful calibration of the imaging system and its luminance generation in particular. If the clinician desires to use such an image analysis in a clinical examination, it is important to perform a calibration known as gamma correction which can correct for the nonlinear relation between image values and intensity on the screen surface[26,27].

Spatial. Just as the mosaic of receptors over the retina indicate that any viewed scene is sampled in the spatial dimension, so too do imaging systems discretely sample any scene they are aimed towards. However, there is one major difference between how the visual system and imaging systems carry out their respective spatial sampling that is important to consider. The visual system is able to operate with one small but high bandwidth region that has a high level of resolution. This region is the fovea. To maximize the utility of the fovea, a system of eye and head movements has been evolved to easily and quickly move the fovea to the object that needs a detailed inspection. The movement of the fovea is so effortless that most of us are rarely aware of our eye movements[28]. However, an imaging system cannot take advantage of the same technique to reduce the bandwidth of the information being processed. The imaging system does not and generally cannot know which part of an image the observer is examining, i.e., which part of the image falls on the observer's fovea. Furthermore, if there are two or more people viewing an image they will often be examining different parts of that image, making it necessary to render all portions of the image at the highest resolution needed. The result is a greatly increased bandwidth for transmitting, processing and displaying of the image by all levels of the imaging system which tends to decrease both the displayed size of the image, pixel width and height, and the extent of the scene covered. As seen in Figure 10, this lack of image size creates a situation where illusions of size and depth are more likely. In laryngeal imaging, the situation is even more problematic because of the inability to move the endoscope around to disambiguate the visual information.

The problem will be reduced as the image processing and storage capabilities of these systems increase. In the meantime, it is recommended that images are displayed showing the full resolution possible. Moreover, if multiple diagnostic techniques are used problems with any one method may be overcome. The greater imaging capabilities of rigid endoscopes are also to be preferred to the flexible endoscope, although the latter has other advantages (e.g., more natural head positioning and the ability to talk).

Chromatic. Whereas the ability to use only three primary colors to reproduce a wide range of colors results from the fact that the human visual system has three cone classes, in practice, it is not possible to choose three primary colors that will reproduce all possible colors. In Figure 4, the black triangle in the middle of the CIE diagram illustrates a representative range of colors that can be reproduced on a good quality CRT display (a standard television set type of monitor)[20]. The corners of the triangle represent the chromaticity coordinates in the CIE system for the three primary colors. Color mixing works on the CIE diagram just like it does on a color wheel (topologically they are the same). A color mixture between two colors always falls on the line between two colors. Thus, the triangle represents all possible colors that can be reproduced by this sample monitor. This region is called the color gamut. As can be seen, there are many colors that cannot be reproduced on this monitor. Many photographic quality color printers often have as many as 6 primary colors with an additional two black inks to create a wider color gamut. In laryngeal imaging, if diagnostically important color changes lay outside the gamut, these colors will not be seen. As most of the colors are in the red region, an important area of study would be comparing the range of reds produced by different pathologies and maximizing the imaging systems in this region of the color gamut.

Each image display device will have a different color gamut so colors may not appear the same on different monitors and from monitors to printers. Differences in the relationship between image value and intensity (gamma) will also impact the way that colors appear in the reproduced image, further leading to differences in color when an image is viewed on different monitors. Thus, two observers seeing the image on two different screens will see a differently colored image. In most cases the differences will be subtle, but if the diagnostic feature relies on a subtle difference, then the difference in coloring of the two images could lead to different diagnostic conclusions and explain some of the differences seen in the literature on diagnostic reliability[3]. To minimize, but not eliminate entirely, this problem, images should not be shared and viewed in isolation. If it is desirable to have a colleague view an image, a comparison image not showing the pathology should be sent. It is more likely that the differences between the images should be preserved across different systems than absolute color values[20].

Stability

Stability refers to whether the imaging system will reproduce the same image value either at different locations on the monitor surface or at different times on the same monitor. Most systems have adequate stability for gross determinations, but vary in ways that can make subtle discriminations problematic[8,29]. Problems of stability to be discussed here arise in the dimensions of luminance and color.

Luminance. A monitor's luminance will change as the monitor warms-up. CRTs take about 20-30 minutes to reach their maximum luminance output; the warm-up times for LCDs are more variable[26]. Diagnostic reading should not be done on a monitor that has just been turned on; the monitor should be allowed to warm up about 20 minutes to reach a more stable luminance value.

Even after the monitor warms up, luminance varies across the monitor surface. As noted by several researchers, the luminance level generally falls off towards the edges of CRT screens, but not in an easily predictable manner[8, 30, 31]. The luminance level is generally more even on LCD flat-panel screens[29]. To minimize the influence of monitor screen luminance, there are sophisticated calibration techniques that can be used[30,31], but the easiest solution is to limit the image to the center of the screen[8]. It is important to note that changing brightness, contrast and other settings on your monitor will undo the utility of any calibration effort. So, once a monitor is calibrated, the monitor settings need to be fixed.

Room lighting also impacts both the luminance level and contrast on the screen. CRTs reflect about 10% of the light hitting the surface of the monitor; LCDs reflect about 1%[32]. This reflected light adds to both the background light level and the foreground light level coming off of the surface of the monitor. This additional light, while making the monitor more intense overall, reduces the contrast in the image; i.e., it washes out the image. Since human perception is driven by being able to detect differences in light levels, e.g., the example of simultaneous contrast in Figure 3, this reduction in contrast can reduce the ability to perceive differences in the image. The problem is far worse for CRTs than LCDs. Figure 11 illustrates the problem. Figure 11a and 11c show the same picture photographed off of a CRT (a) and LCD (c) screen in a dark room. In Figures 11b and 11d, the same screens are photographed with the room lights on. Notice how the reflected light, particularly on Figure 11b, causes the image to wash out. The washing out of the images can even be seen in a simple analysis of the images. The range of luminance values in the dark image (11a) is far greater (29 to 255) than for the room light image (64 to 246)(11b). While the camera is not the eye, this measure does indicate the loss of contrast in the image in room light. This loss of contrast will show up in the loss of details and fine distinctions which maybe necessary for early diagnosis of some disorders. There is far less difference in how the image appears on the LCD whether in the dark, in 11c, and with the room lights on, in 11d. If the room has an open window that lets sunlight fall on the monitor screen, the situation is far worse than illustrated in Figure 11. In terms of laryngeal imaging, when the monitor is a CRT, it is important to have the room lights as low as possible during the examination of the images. For LCD screens, the issue of room lighting is far less acute, though windowed rooms should be avoided[32].

Chromatic. Even as different monitors will generate different light levels and the light levels vary across the screen surface, these differences apply to the three primary colors differently[8]. These variations between primary colors can cause a shift in the color that is being generated. Thus the colors can shift, usually slightly, from one viewing to the next and as the image is moved around the screen. With LCDs, the color can also shift with the direction of viewing, though the issue of viewing angle on LCD screens has become far less important as the technology has matured.

Lighting also affects the colors on the screen. Since room light is generally white, the effect of room light and sunlight is to desaturate the colors on the screen. In fact, in the original color images for Figures 11a and 11b from the CRT, the effect on color is even greater than the effect on luminance. This is another reason to establish standards for lighting when examining images for clinical purposes.

Conclusions and Recommendations

A sound understanding of the visual system, imaging systems and how they interact can lead to a more critical comprehension of laryngeal imaging systems and their limitations. The key issues may be summarized as follows. First, the visual system (a) actively seeks information, (b) organizes the input to aid interpretation, and (c) follows rules to how the visual system organizes sensory input. In general, the pattern perceived is the simplest and most stable interpretation of the visual scene. Laryngeal imaging involves both the imaging system and the visual system. The visual system organizes the incomplete information provided by the imaging system into a coherent pattern which may or may not accurately or even informatively represent the original scene.

Second, the visual system adapts. The visual system can change its mode of operation to optimize its performance in different environments. The adaptability of the visual system allows it to operate over a larger range of conditions and can help overcome some of the stability problems of imaging systems. However, this ability to adjust may obscure changes in imaging system operation that are important, such as the dimming of the bulb on the endoscope which will reduce image contrast and brightness.

Third, vision is contextual, e.g, Figures 3 and 4. Even the brightness of any region depends as much upon the luminance of surrounding areas as the luminance of that region. The breadth of the visual scene that is experienced is an important aid in interpreting any scene. A restricted field of view limits the visual system's ability to decipher ambiguous stimuli. In the laryngeal imaging, the clinician has a very limited field of view. Another important source of context is past viewing of similar scenes.

Fourth, the imaging systems sample their images. The sampling of luminance, spatial and chromatic dimensions have been discussed. The sampling can both lead to loss of important information and illusions. The range of sampling of crucial features should be studied.

Fifth, imaging systems are not entirely stable. The stability can be both inherent to the system or the result of the viewing environment.

Together these issues leads to an important recommendation: calibrate the clinician's visual system by regular examination of a broad range of both normal and abnormal laryngeal images[3,4]. The visual system is particularly sensitive to change and to the context. Therefore, imaging systems will better represent differences in images than in its absolute values. Thus, viewing a range of images on a given system will help the clinician to see how any given system will depict any given pathology at any given time.

Several corollary recommendations result from this main recommendation. When sharing images, it is important to always share at least two images, one with the putative pathology and one without. It also seems wise never to rely on any single method to examine laryngeal function. By using a range of techniques, both imaging and non-imaging, all of the information obtained creates a rich context which can help disambiguate information provided by any one method of examination. Given that limitations are inherent in any method of examination, it is important to carefully consider all of the information and look at the total picture[6].

In addition to calibrating the clinician's way of viewing laryngeal images it is important to set standards for monitor use and calibrate the imaging system. Monitors differ from each other; any given monitor will vary in its functioning from one time to another. To increase the stability of a monitor, it is recommended to let is warm up for 20 minutes. The lighting environment also influences the contrast and color on a monitor. Imaging devices need standard and regular calibration. However, it seems unlikely that most clinicians will develop the skills or have the time or patience to do the extensive calibrations required. It seems more prudent to work with manufacturers to develop calibration protocols that can be done as a part of regular servicing of the equipment. The calibration should include all facets of the imaging system from endoscope to monitor. It is important to get the imaging system into a standard condition of operation that is comparable from one location to another. With regards to the lighting, it is important to develop standards for the lighting of the room while observing images making the viewing environment standard as well. This recommendation about lighting is more important for CRTs than LCDs under normal room lighting conditions.

Laryngeal imaging has played an important role in advancing the care of the voice[1,2,3,4]. However, as with any technology, it has its limitations and problems. Understanding how the visual system operates within the limits of imaging systems is important. This understanding will help improve the use of information from laryngeal imaging and also develop criteria for the best use of this equipment. By paying more attention to the visual system and the nature of imaging systems, the impact of laryngeal imaging on the care of the voice can be even greater.

References

[1] Hertegård S. What have we learned about laryngeal physiology from high-speed digital videoendoscopy? *Curr Opin Otolaryngo Head Neck Surg* 2005; 13: 152-156.

[2] Solomon NP. We Have to Look Where? A Historical Perspective on Laryngeal Imaging. Presented at the 35th Annual Voice Foundation Symposium: The Care of the Voice, 2006; Philedalphia, PA.

[3] Milstein, CF. A Current Perspective on the Role of Visual Perception in Voice and Medicine. Presented at the 35th Annual Voice Foundation Symposium: The Care of the Voice, 2006; Philedalphia, PA.

[4] Bless DM. Seeing and believing: The limits and joys of visual-perceptual laryngeal assessment. Presented at the 35th Annual Voice Foundation Symposium: The Care of the Voice, 2006; Philedalphia, PA.

[5] Hicks DM, Ours TM, Abelson TI, Vaezi MF, Richter JE. The prevalence of hypopharynx findings association with gastroesophageal reflux in normal volunteers. *J Voice* 2002 Dec; 16(4): 564-579.

[6] Kahneman D. Judgment and decision making: A personal view. *Psychol Sci* 1991; 2(3): 142-145.

[7] Stegeman, C. The art of Observation: Visual thinking for the healthcare professional. Presented at the 35th Annual Voice Foundation Symposium: The Care of the Voice, 2006; Philedalphia, PA.

[8] Krantz JH. Tell me, what did you see? The stimulus on computers. *Beh Res Meth Inst Comp* 2000; 32: 221-229.

[9] Barlow HB. Summation and inhibition in the frog's retina. *J Physiol* 1953; 119: 69-88.

[10] Krantz JH. Computerized Tutorials: Animation vs. Modeling. Presented at the 34th Annual Meeting of the Society for Computers in Psychology (SCiP), 2004, Minneapolis, MN.

[11] Daniel PM, Whitteridge D. The representation of the visual field on the cerebral cortex in monkeys. *J Physiol* 1961; 159: 203-221.

[12] Osterberg G. Topography of the layer of rods and cones in the human retina. *Acta Ophthal* 1935; (Suppl. 6)U.

[13] Proffitt DR. Embodied perception and the economy of action. *Persp Psychol Sci* 2006; 1(2): 110-122.

[14] Kanizsa G. Subjective contours. *Sci Am* 1976; 234(4): 48-52.

[15] Gerrits HJMJ, Vendrik AJHJ. Simultaneous contrast, filling-in process and information processing in man's visual system. *Exp. Brain Res* 1970; 11: 411-430.

[16] Lamy D, Segal, H, Ruderman, L. Grouping does not require attention. *Percep Psychophys* 2006; 68(1): 17-31.

[17] Abrams RA, Christ SE. Motion onset captures attention: A rejoinder to Franconeri and Simmons (2005). *Percep Psychophys* 2006; 68(1): 114-117.

[18] Farah, MJ. Is face recognition "special"? Evidence from neuropsychology. *Behav Brain Res* 1996; 76: 181-189.

[19] Jackson MC, Raymond JE. The role of attention and familiarity in face identification. *Percep Psychophys* 2006; 68(4): 543-557.

[20] Silverstein LM, Merrifield RM. *The development and evaluation of color systems for airborne applications* (DOT/FAA Tech. Rep. No. DOT/FAA/PM-85-19). Washington, DC: DOT/FAA: 1985.

[21] Boynton RM, Fordon J *Bezold-Brücke* hue shift measured by color naming technique. *JOSA* 1965; 55(1): 78-86.

[22] Doughtery R & Wade, A Vischeck (URL: http://www.vischeck.com/ Downloaded July 25, 2006).

[23] Proctor RW, Van Zand T. *Human factors in simple and complex systems*. Needham Heights, MA: Allyn & Bacon; 1994.

[24] Goldstein EB. *Sensation and perception*, 6th Ed. Pacific Grove, CA: Wadsworth; 2002.

[25] Westheimer, G. The Ferrier lecture, 1992. Seeing depth with two eyes: stereopsis. *Proceedings: Biological Sciences* 1994; 257: 205-214.

[26] Metha AB, Vingrys AJ, Badcock DR. Calibration of a color monitor for visual psychophysics. *Beh Res Meth Instr Comp* 1993; 25: 371-383.

[27] Olds ES, Cowan WB, Jolicoeur, P Effective color CRT calibration techniques for perception research. *JOSA A* 1999; 16: 1501-1505.

[28] Volkmann FC, Riggs LA, White KD, Moore RK. Contrast sensitivity during saccadic eye movements. *Vision Res*. 1978;18(9):1193-9

[29] Krantz JH. Stimulus delivery on the web: What can be presented when calibration isn't possible. In: Reips Bosnjak, Eds. *Dimensions of Internet Science* Berlin: Pabst Science Publishers; 2001: 113-130.

[30] Cook JN, Sample PA, Weinreb RN. Solution to spatial inhomogeneity on video monitors. *Col Res Appl* 1993; 18: 334-340.

[31] Hu QJ, Klein SA. A two-dimensional lookup table to correct the spatial nonlinearity on CRT displays. *Soc for Infor Disp Dig of Tech Papers* 1994; 25: 19-22.

[32] Krantz JH, Silverstein LD, Yeh Y-Y Visibility of transmissive liquid crystal displays under dynamic lighting conditions. *Hum Fac* 1992; 34: 615-632.

Figure Captions

Figure 1. A sample laryngeal image. Image by Kay Pentax, Permission to use granted.

Figure 2. The Kanizsa triangle showing illusory contours.

Figure 3a. Simultaneous Contrast. The two central squares are exactly the same gray.

Figure 3b. Simultaneous Contrast in a laryngeal image. The area indicated by the 1 is actually slightly less intense on average than the area indicated by the 2. Origninal Image by Kay Pentax, Permission to use granted

Figure 4. The 1931 CIE Diagram.

Figure 5. The first stages of color vision processing by the visual system. The three classes of cones are on the left side. They connect, most importantly for the present purpose, to two types of color opponent cells. These are the Blue-Yellow (B/Y) cells and the Red-Green (R/G) cells. The teal colored connections are excitatory and the dark grey connections are inhibitory.

Figure 6. A montage of stroboscopic images of a healthy larynx. (a) How it appears to a color normal. (b) How it appears to a with protanopics vision. (c) How it appears to a person with deuteranopic vision. (d) How it appears to a person with tritanopic vision.

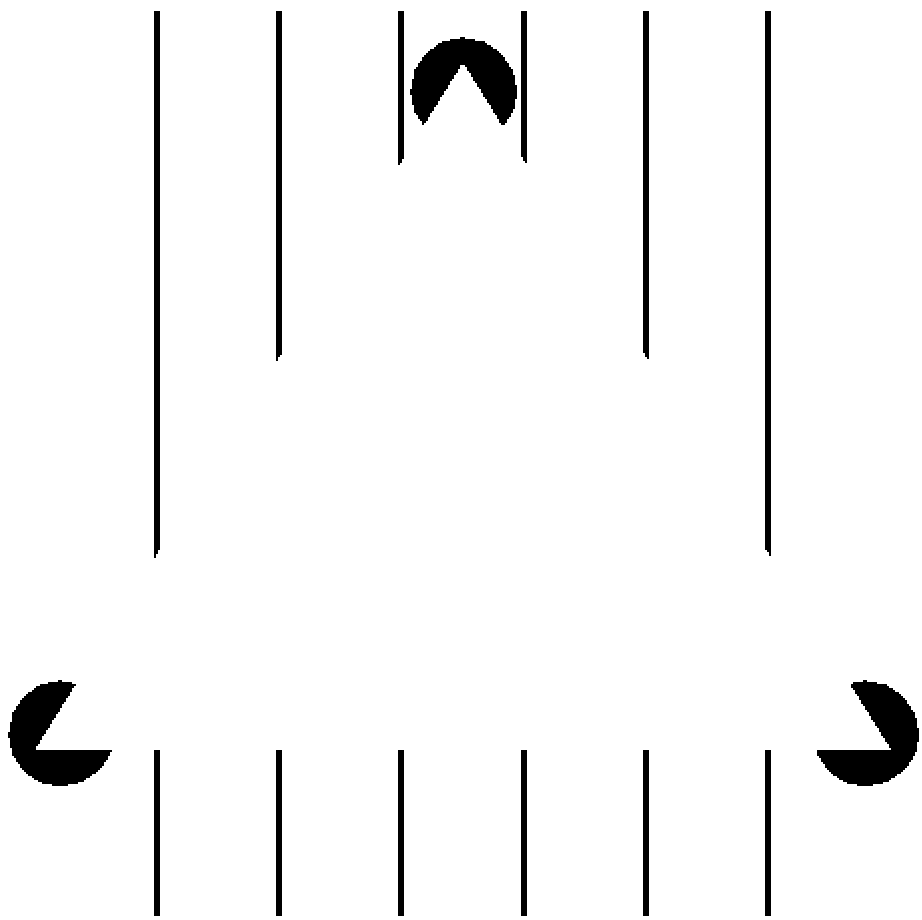
Figure 7. A description of the wagon wheel illusion.

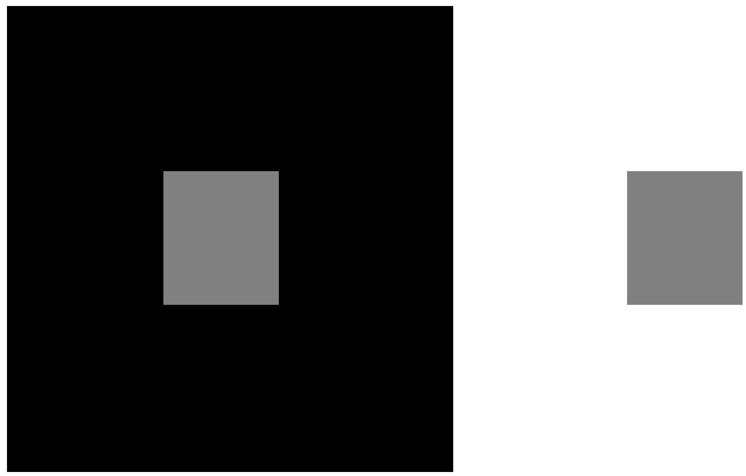
Figure 8. An illustration of pictorial monocular depth cues.

Figure 9. An illustration of how binocular depth cues of vergence and disparity work in the visual system.

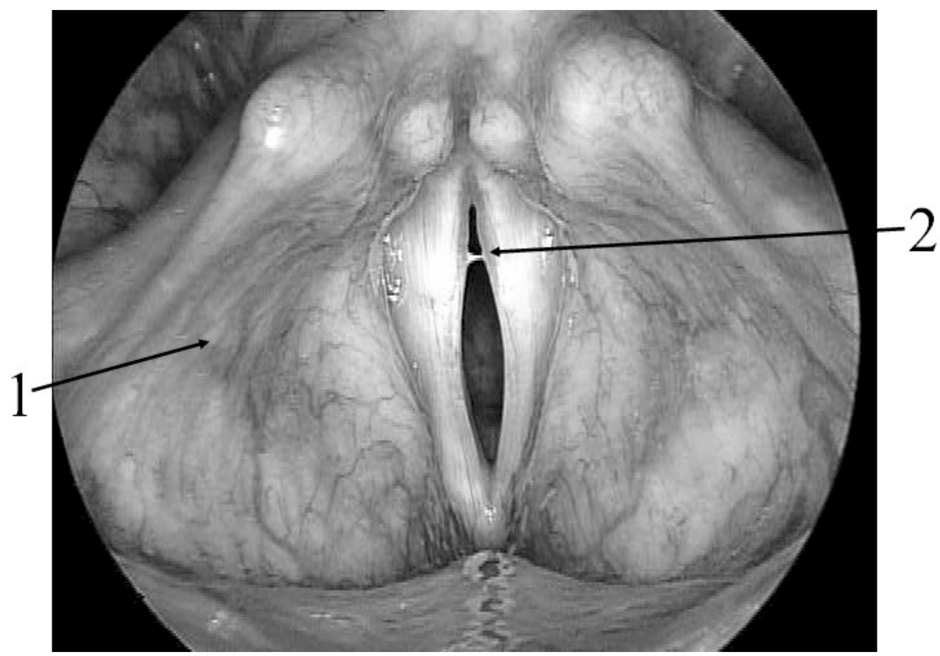
Figure 10. These two mugs are exactly the same size.

Figure 11. The effect of lighting on CRT and LCD monitors. (a) CRT monitor in the dark. (b) CRT monitor in room light. (c) LCD monitor in the dark. (d) LCD monitor in room light.

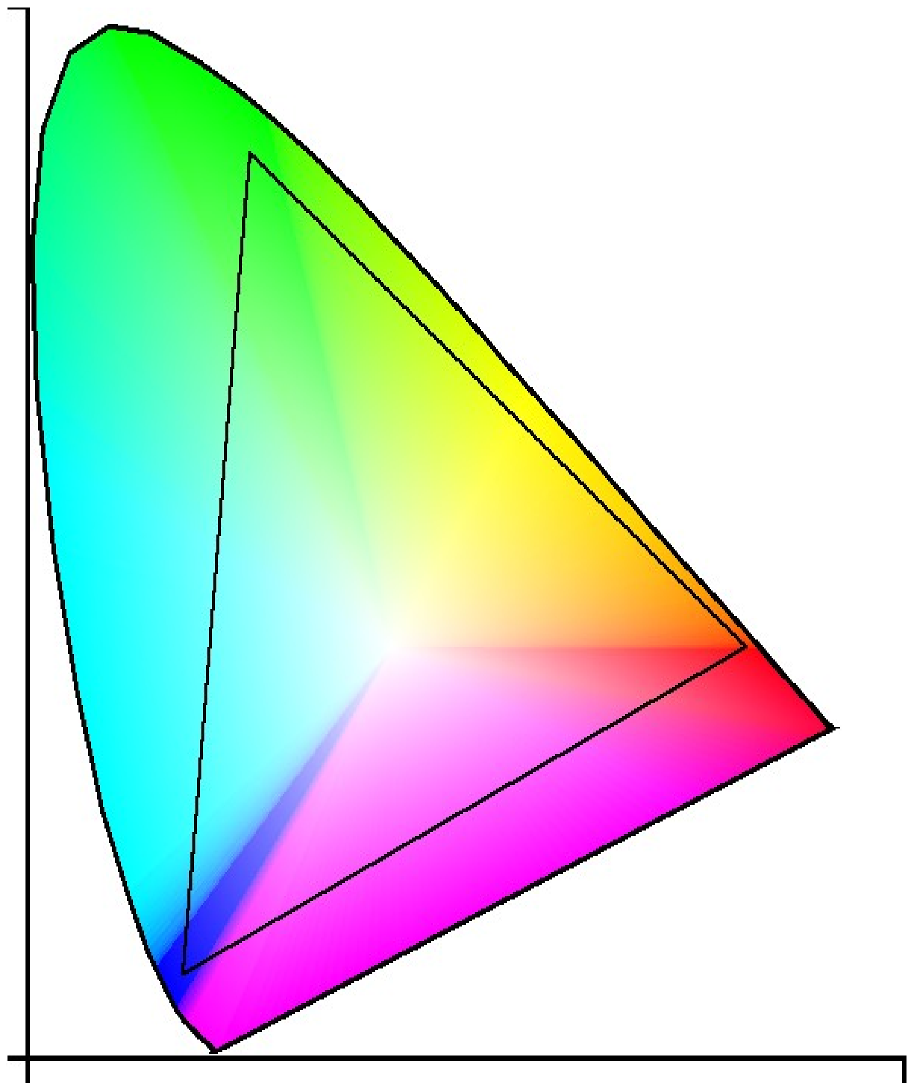


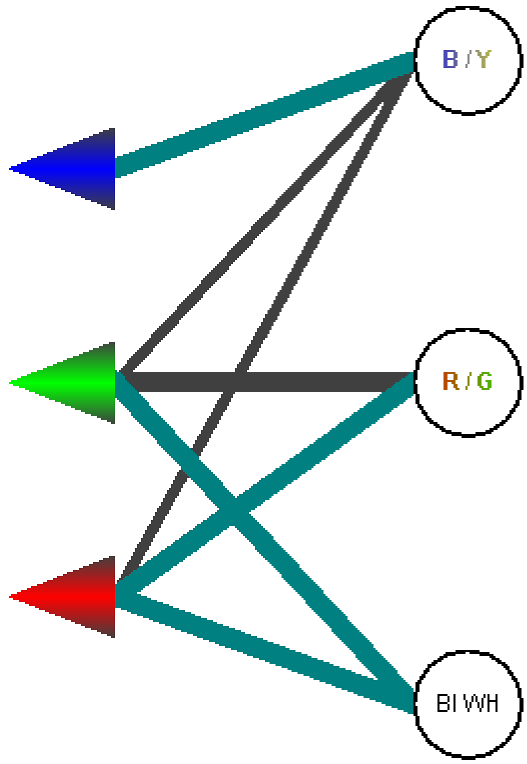


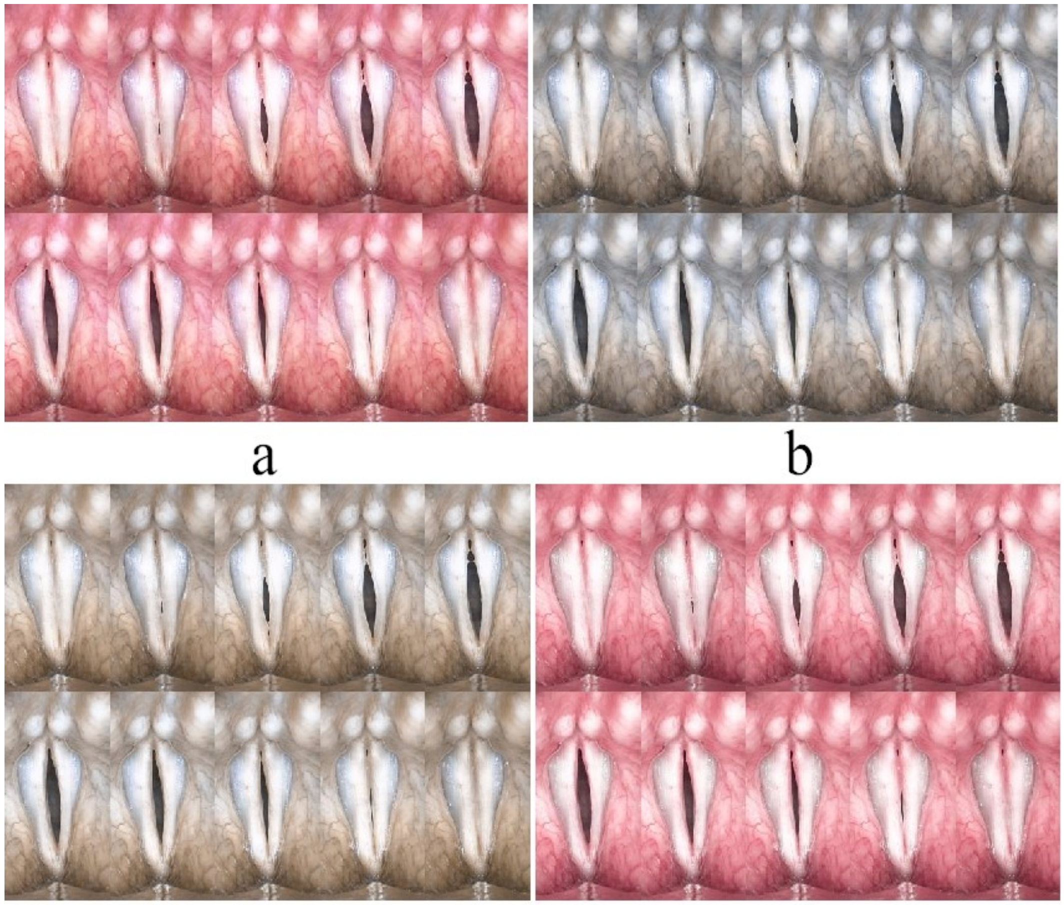
(a)



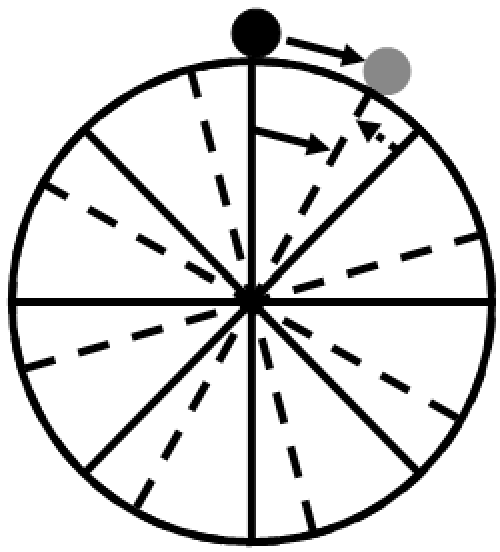
(b)

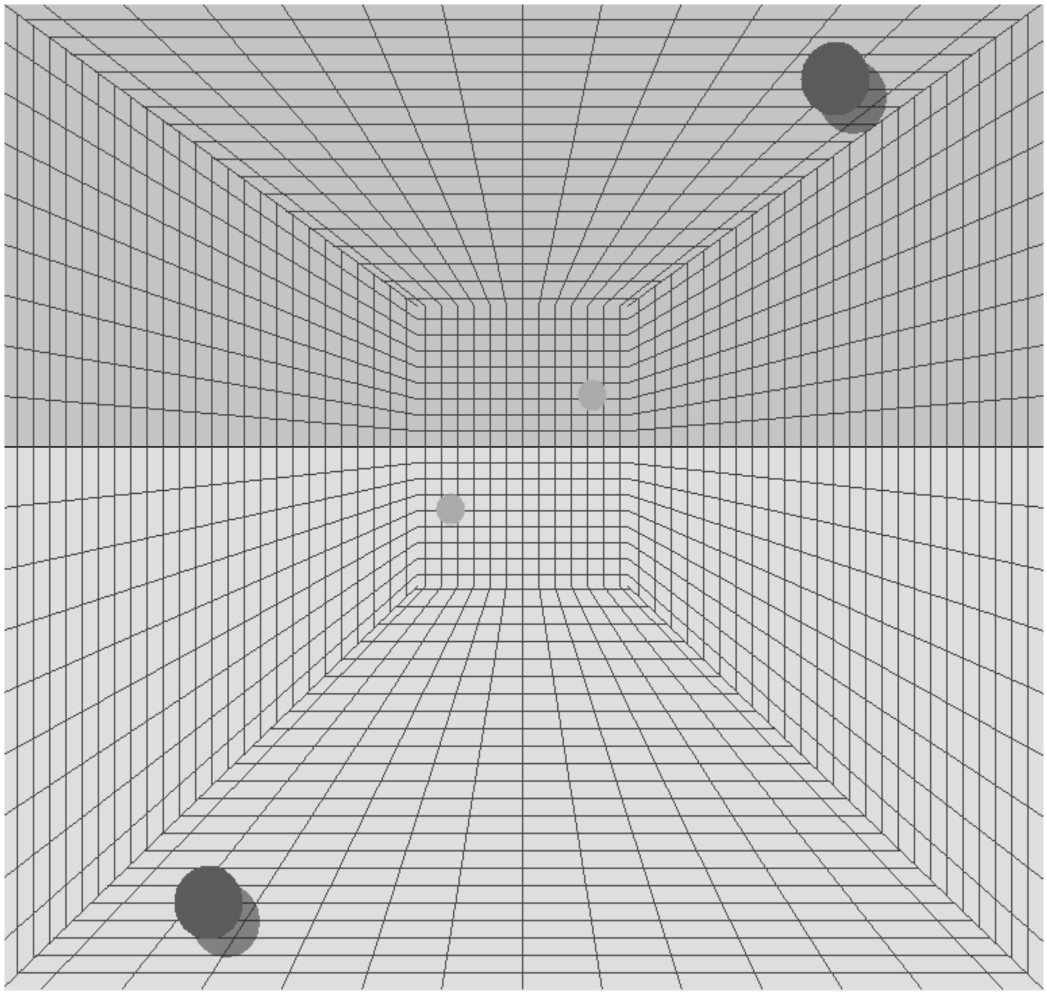


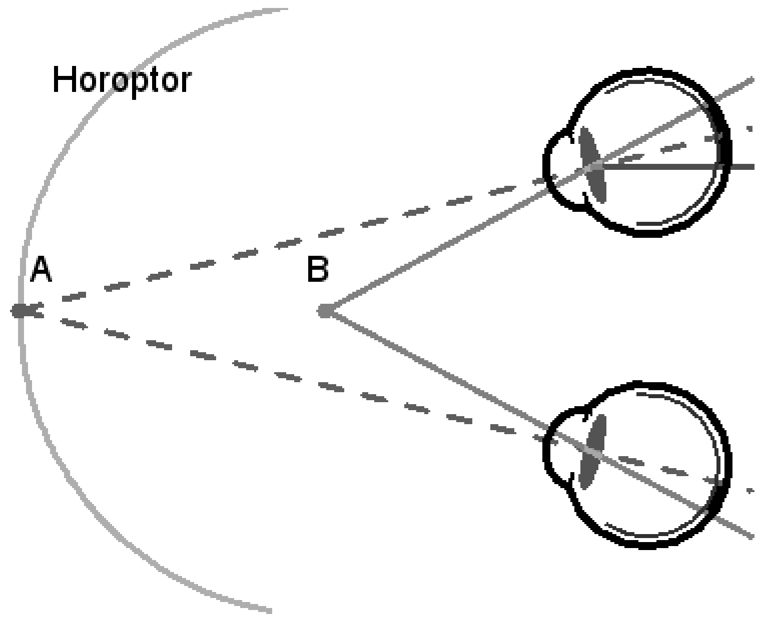

B / Y
R / G
Bl Wh



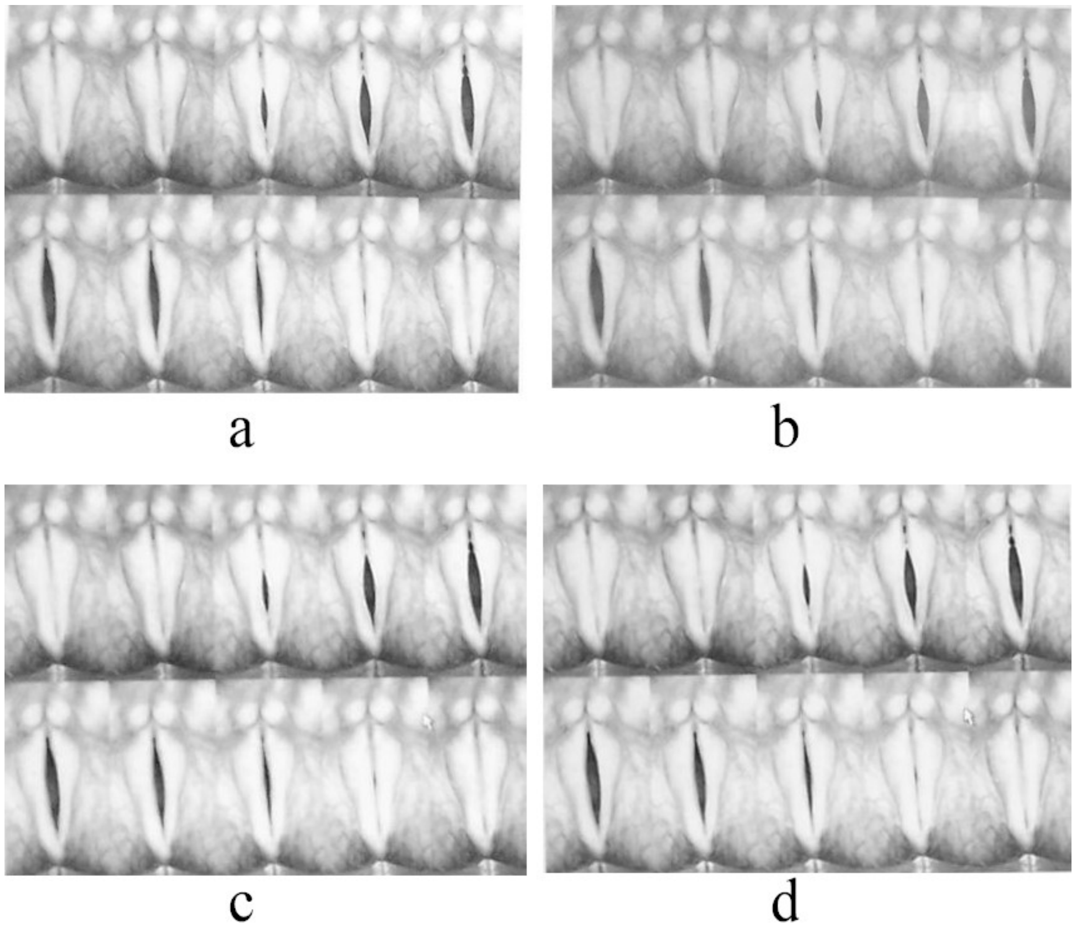
a b c d






Horoptor
A
B


LoonMagic™


a
b
c
d