Chapter 4
Behavior: Basic Psychology of the User

Abstract This chapter examines what are described as user behavioral characteristics. These are characteristics that are related to perception in broad terms. The chapter starts by defining some behavioral terms and concepts that are used in this and subsequent chapters. We then describe in detail several aspects of the two main perceptual systems that are involved in interacting with computer-based systems: vision and hearing. For each of these aspects we consider some of the implications they have for system design. We finish by introducing the topic of motivation to help explain why individual users may behave in a particular way when carrying out a task.

4.1 Introduction

When we refer to behavioral characteristics we are really talking about things that are linked to sensation and perception, in general terms. We know that people have five basic senses: sight, hearing, touch, smell, and taste. Sensation occurs when the sense organs (eyes, ears, and so on) are stimulated, and they generate some form of coding of the stimuli. Perception occurs when this coded information is further interpreted using knowledge of the current context (physical, physiological, psychological, and so on) to add meaning. The process of perception is subjective: simply presenting designed stimuli in such a way that they will be sensed accurately does not necessarily mean that they will be perceived in the way that the designer intended.

Sight, hearing, and smell all sense stimuli that appear at some distance from the body. In other words they detect distant (distal) stimuli. Touch and taste, however, rely on contact being made with the stimuli. This means that the stimuli have to be very close to the body, which is why they are described as proximal stimuli.

We need to think about the user's behavioral characteristics when designing systems because we want to make sure that the system fits the user. A well-designed system will take into account the user's ability to detect changes in an interface, for example, and we can use this information to inform the design of screen layouts to

help the user make the most of that ability. It is also important to remember that these abilities will vary across users, tasks, and contexts. If your system will be deployed in a brightly lit office, for example, then reflected light and glare are likely to affect the user's ability to read what is displayed on the screen.

In most computer-based systems we tend to privilege vision and hearing over the other senses, although the rapid uptake of smartphones and tablet computers has increased the importance of considering touch (described in Chap. 3). Often it is important to consider how the different sensory modalities can be used together to provide further useful information to the user. In dark, or dimly lit conditions, for example, using different shaped knobs and switches can be used to exploit the sense of touch to impart information to the user. Similarly, if a user has impaired vision, we may want to present information in such a way that it can be exploited by other senses, such as touch and hearing.

Once we have an understanding of how people sense and perceive things, and how the different senses work together, we can start to think about broader design questions such as:

• What is the best way to present information to the user?
• Can the user detect the information that we are presenting to them?
• If we are presenting a lot of information to the user, should we be presenting it using more than one sensory channel (e.g., visual and auditory channels are often both used for alarm information).

Next, we introduce some basic concepts from behavioral psychology that will help you to understand and interpret the rest of the chapter. We then provide quite detailed descriptions of the visual and auditory perceptual systems. There are many aspects of the visual and auditory systems that need to be considered when designing systems. For this reason, we consider the implications for system design at the end of each of the sections in this chapter. Note that we omit the senses of smell and taste here because their use is almost exclusively limited to systems that are dedicated to tasks that only involve detecting particular smells and tastes.

4.2 Behavioral Psychology Terminology

Within psychology many terms have been developed to describe different aspects of behavior, extending far beyond perception. Here we introduce some terms that may help you when thinking about the user and usability with respect to perception.

4.2.1 Thresholds and Just Noticeable Differences (JNDs)

Each sense has a threshold for detecting stimuli. Sounds that are too quiet or visual stimuli that are too faint cannot be sensed. Table 4.1 gives some representative

Table 4.1 Some human sensory thresholds (under ideal conditions)

Sight	A candle flame seen from 50 km on a clear dark night (100 quanta to the eye, or 10 quanta absorbed by the rods)
Sound	The tick of a watch from 6 m in very quiet conditions (0.0002 dynes/cm^2)
Taste	One gram of table salt in 500 L of water (0.0001 M)
Smell	One drop of perfume diffused throughout a three room apartment or 1×10^{-12} mol/L of ethyl merchantman
Touch	The wing of a bee falling on your cheek from a height of 1 cm (10 mg force)

thresholds under ideal conditions for the various senses (taken from Galanter 1962). It should be noted that thresholds can vary across tasks and contexts. In a room full of noisy machinery, for example, you may need to increase the volume of alarm warning sounds, so that they can be heard by users.

The sensitivity of perception can be measured, based on how small a change can be detected. This difference is called a *just noticeable difference*, or JND. In vision, for example, it is how much brighter a scene has to be in order for someone to report it as being brighter; and in hearing it is how much louder a sound must be to be noticeably different. Although JNDs are objectively measured, their magnitude is subjective, varying across users, tasks, contexts, and modality.

4.2.2 Habituation

All living things react to stimuli. If a stimulus occurs repeatedly and is not regarded as salient they will habituate to it. People who live beside a railroad track, for example, grow accustomed to the noise of the trains to a point where they no longer notice it. In other words, people learn which stimuli are not salient (in the current context), and hence do not require further processing. Habituation effectively frees up cognitive resources, which allows people to use those resources to deal with new stimuli as they are presented. Constant "confirm action" boxes, for example, will become habituated to.

4.2.3 Signal Detection Theory (SDT)

Accuracy is an important aspect of performance. In many cases, simple measures of correctness, such as the number of targets that were recognized and the number that were missed, are sufficient. This is too simple a measure, however, when examining performance under conditions where the ability to be correct is difficult, and the types of mistakes are important. Table 4.2 provides a way of summarizing this more complex situation: if you report seeing something when it is there, it is called a hit; if you fail to report seeing something when it is there, it is called a miss. Similarly,

Table 4.2 Types of responses to a signal

Response	Signal present	
	Yes	No
Yes	Hit	False alarm (FA)
No	Miss	Correct rejection (CR)

reporting seeing nothing where there is nothing there is called a correct rejection, whilst reporting seeing something when it is not there is called a false alarm.

Human visual behavior often includes searching, scanning, and monitoring. Vigilance tasks, which are a subset of these tasks, usually involve an extended time period between the signals that the user is supposed to recognize. While it should be possible to moderate the performance of these tasks based on the dimensions used in the table, it is often useful to analyze these situations using Signal Detection Theory (SDT) (Swets 1973; Swets et al. 1961; also see Wickens et al. 2014 or Wickens 2002).

SDT was developed to explain the task of identifying enemy planes on hard to read radar displays. It has been applied to a wide range of classification and decision making tasks. Swets (1973), for example, used SDT to summarize how well doctors could classify cancer tumors from X-ray films. It can also be applied within computer games to explain the potential problems when players search for enemies or resources.

Figure 4.1 shows how the key parameters in SDT relate to the ability to distinguish the object of interest (signal) from distracting items (noise). The set of signals is assumed to be normally distributed about a point some distance away from 0, whilst the set of noise is normally distributed around 0 (indicating no valid signal strength). d' represents the distance between the mean of the noise and the mean of the signal, and is an inherent property of an observer and stimuli. Observers set a threshold (here, beta, β, some authors call this lambda, λ, or yet another Greek character) as a parameter that is inherent in the observer. Observations above the threshold are classified by the observer as signal and observations below are classified as noise. Thus, the parts of the signal above the threshold are hits and the parts of the noise distribution are false alarms (FA). Observations below the threshold are either correct rejections (CRs) if they are noise or misses if they were part of the signal.

Observers can often adjust the threshold to take account of the relative costs of the four responses in Table 4.2. The cost of false alarms and misses will influence where the threshold is set. Where misses are costly, the threshold will be to the left, classifying more signal and more noise as positive responses. Where false alarms are expensive compared to misses, the threshold will be moved to the right, eliminating more of the noise but also part of the signal.

Signal detection theory is often described in terms of sensitivity and bias. The sensitivity (which is the same as d') refers to the separation between the signal and noise distributions. Where the separation is small, and the two distributions
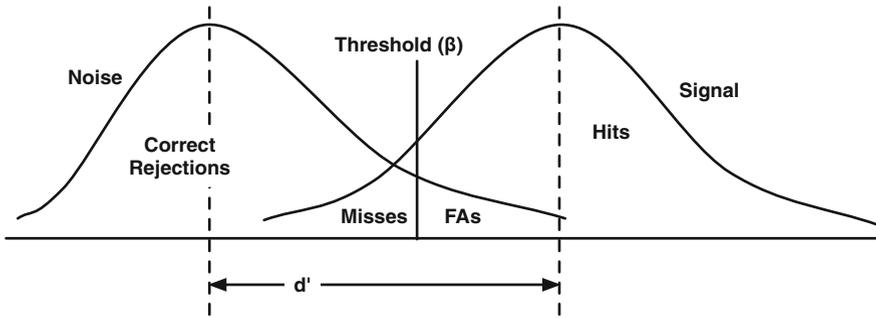
Fig. 4.1 Signal and noise distributions in signal detection theory

overlap considerably, sensitivity is said to be low. Where the separation is large, sensitivity is said to be high. The response bias (which is the same as β) describes how conservative the user is in responding. If the user responds conservatively, this is because a lot of evidence is needed to be sure that what has been observed is really a signal (a true hit), and not just noise.

4.2.4 Implications for System Design

It is important to be aware of the context in which your system will be used. A data indicator on a display, for example, will be sensed differently in a brightly lit location and in a dimly lit location. Users also find it hard to differentiate between elements in an interface that use stimuli that are separated by less than a JND (such as two shades of the same color). If it is important that some elements on a display which are visually similar (such as same shape, but slightly different color) are processed differently, they should be made distinct by separating one or more dimensions of their appearance by several JNDs (several shades of color, for example).

If a system generates lots of false alarms or interrupts, users may habituate to the alarm sounds, like in Aesop's Fable about the boy who cried wolf. Then, when a real emergency happens, users may ignore the alarm. It is therefore crucial to minimize the number of false alarms. Pay close attention to how and where alarm limits are set, and make sure that you take into account how noisy the data values being judged are. This may involve preventing the alarm from oscillating on and off when there are recurrent spikes in the values that are due to artifacts such as measurement error.

For systems where the users have to perform vigilance tasks (like monitoring the screens of hand luggage scanners at airports to identify dangerous and prohibited items), you will need to calculate the costs of false alarms and misses. If the costs are high, there are some measures that you can employ to increase the users' sensitivity, such as showing target examples that they can refer to for comparison, and to

change their response bias, such as making clear to them the costs of false alarms and misses (see Wickens et al. 2014 and Wickens 2002 for more complete lists).

4.3 The Physiology of Vision

In the next few sections we examine the details of vision. We start by offering a high-level description of the physical structure of the eye. This should help you understand the basic idea of how vision works and what some of its important limitations are. Unlike a camera snapshot, for example, the eye does not capture everything in a scene equally, but selectively picks out salient objects and features from the current context, and focuses on them so they can be processed in more detail.

4.3.1 Overview of Vision

For normally sighted people, vision is by far the most widely used sense. Vision is important in everyday work because it allows us to use interfaces like those shown in Fig. 4.2. Understanding the basics of how human vision works, including its strengths and weaknesses, will help you to design systems that more closely match your user's visual capabilities.

4.3.2 The Basic Structure of the Eye

Figure 4.3 shows the basic structure of the eye. The important physiological features that you need to be aware of are the lens, the retina, the rod and cone cells (the sensory receptors in the eye which respond to different light waves), the fovea, and the optic nerve.

 When you look at an object, an image of it is projected onto the eye. The angle that is subtended by that object at the eye is described as the *visual angle*. The ability to discriminate between two objects that are close together is described as *visual acuity*. This is usually expressed in terms of the minimum visual angle that can be resolved, for which the standard is normally a gap subtending 1 min of arc (1/60 of a degree). Acuity is usually expressed as a fraction which expresses the ratio of the standard distance used in eye-tests (20 ft or 6 m) to the distance at which a gap subtends 1 min of arc. So when you hear someone described as having 20/20 (or 6/6) vision, the numbers refer to their visual acuity.

 The light coming from an object is focused by the lens and projected onto the retina at the back of the eye. Muscles attached to the sides of the lens contract in order to thicken the lens and bend the light more to achieve a clear focus when the

Fig. 4.2 An interface that has to be recognized quickly and accurately
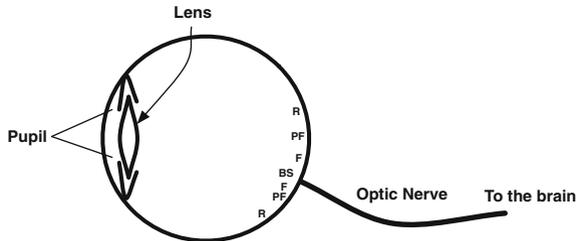

Fig. 4.3 Basic structure of the eye (not to scale). The retina is the area (*roughly circular*) from R to R. The parafovea is from PF to PF. The fovea is from F to F. The blind spot is at BS, where the optic nerve leaves the eye and blood comes in

stimulus is near. These muscles relax when the object is further away. This process of contracting and relaxing is called accommodation. As people age, the lens stiffens, which makes it harder for the muscles to affect its curvature as much, so glasses are prescribed to compensate for the reduced level of accommodation.

The *fovea* constitutes a small part of the retina, about 1–2° of visual arc, which approximately equates to the angle covered by your thumbnail when viewed at arm's length. The receptor systems that permit visual acuity are concentrated only in the fovea, so we need to fixate on an object to have a clear image of it. Surrounding the fovea is the parafovea, which has a lower level of visual acuity than the fovea.

You can demonstrate the existence of the fovea to yourself by staring at a dot in the center of a page. Only the area immediately surrounding it will appear in clear focus, so to clearly perceive a scene of any size the eyes must move around. These semi-conscious movements are called saccades and take approximately 200 ms to

program, and last 20–200 ms, which is just about the interval of the persistence of vision.

The *pupil* of the eye reacts to the amount of light falling on the retina by expanding and contracting in such a way as to keep that amount approximately constant. This feedback mechanism operates within a certain fairly narrow range of illumination (about 16–1), but the enormously greater variation in retinal illumination (about 10^9–1) demands an intermediate mechanism to allow the eye to function over the whole range of illumination. This mechanism is called *adaptation*. Adaptation is one of the most profound and pervasive sensory phenomena. Our eyes are prevented from adapting to what they are seeing by continually moving. These normal movements, which are normally unnoticeable, are faster and smaller than saccades, and described as micro-saccades.

You may have had the experience of entering a theatre from the bright outdoors and stumbling to your seat in the darkness, tripping over the feet of people already seated. In a few moments you will have adapted to the darkness, that is, you have become more accustomed to the lack of light. This mechanism applies to the sense of smell too: when walking into a kitchen where someone is cooking you will notice the odors as very strong, but after a few minutes they become less noticeable. These examples illustrate that after exposure to a stimulus the sensitivity to that stimulus reduces; after removal of the stimulus, the sensitivity returns.

4.3.3 Using Eye-Tracking to Measure Eye Movements

A lot has been learned about how people read and use interfaces by using eye trackers (Duchowski 2007; Holmqvist et al. 2011). Eye trackers are electro-optical systems (typically cameras that can be automatically focused and adjusted) that record where a user has focused their eyes. Simple systems measure where a single eye is looking. Better versions of this simple system can do this without the user having their head fixed, and many can now measure this non-intrusively. Figure 4.4 shows a head-mounted eye-tracker and two example summaries of eye-tracking data. More complex systems can measure both eyes and note how the eyes work together (e.g., strabismus, a measure of how much the eyes have focused onto something near to the head by rotating).

Simple eye-trackers can generally tell which line the user is looking at (to an accuracy of typically 0.5–1° of angular measure, which is about a line of 12 point text on a display at 12 in.). Better eye trackers can tell which letter or which part of the letter the user is focusing on (down to 0.25°).

A useful exercise that will help you to appreciate the accuracy of eye-trackers is to compute the size of a 10 point (1/7 in.) letter at 12 in. using trigonometry. When you have done this, you should see that the tolerance in eye-tracker accuracy is less than the size of the fovea. So although we may know where the center of the user's fovea is focused, there could be several other items that are also projecting onto it and to which the user is attending.
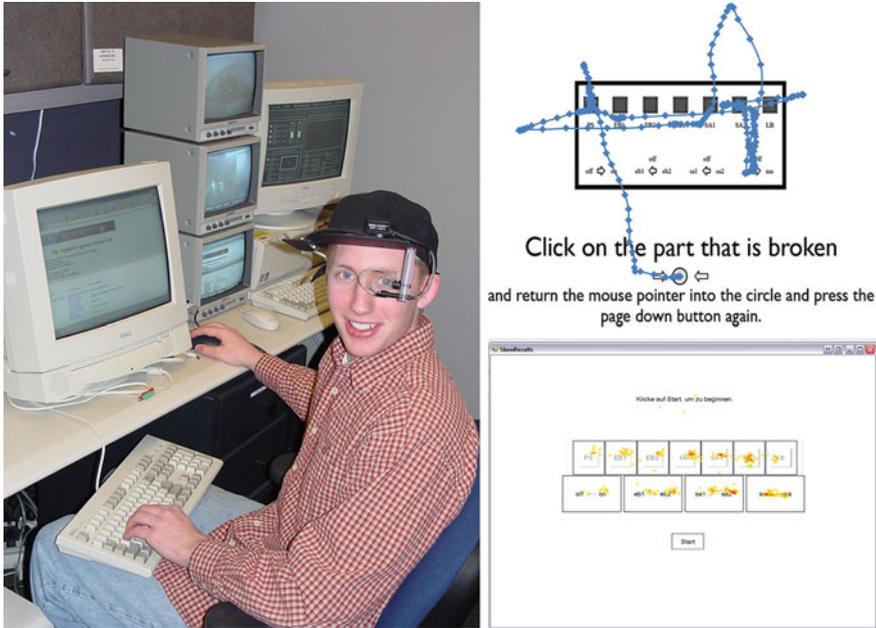
Fig. 4.4 A head-mounted eye-tracker and an example analysis showing the order that a user looked at the interface (*top*) and how long they looked at each part (*bottom*) (photo by Ritter, used with permission of the analyst pictured; analyses courtesy of Maik Friedrich 2008)

In addition to being widely used in psycholinguistics to understand how people read text, eye-tracking is increasingly being used in other fields such as HCI and advertising (Moore and Churchill 2011; Navalpakkam and Churchill in press; Nielsen and Pernice 2010). The availability of relatively cheap, non-intrusive trackers has seen them being used more widely in exploring how people process the text on web pages. With the use of appropriate software you can quickly identify the hot spots on a page, which are areas that the users look at for most of the time.

4.3.4 Rods and Cones

The *retina* at the back of the eyeball is composed of two basic types of cells—*rods* and *cones*—organized in layers. The rods and cones sense light, and mark the start of the transition from sensation to perception with light energy being converted into electrical energy.

According to duplicity theory, the differences between the rods and cones result in two different receptor systems in the eye: one that is best suited to daylight, and one best suited to twilight or moonlight. There are 6 million cones, mostly located in the fovea. These function in daylight conditions and are responsible for color vision. The 120 million rods are distributed across the retina beyond the fovea.
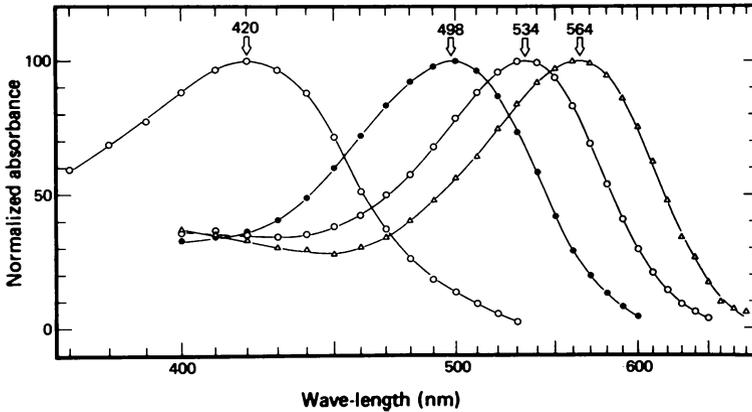
Fig. 4.5 Diagram of sensitivity in nanometers (nm) of rods and cones (reprinted with permission from Bowmaker and Dartnall 1980)

They are much more sensitive to light than the cones and are active in dark conditions; in bright conditions they become overloaded. The distributions of rods and cones overlap in the parafovea. The rods are mostly located beyond the parafovea, in the periphery of the retina, and are very sensitive to movement.

The visual receptors respond to light waves. The term *light* is used to describe electromagnetic energy in the visible range of wavelengths approximately 400–700 nm (see Fig. 4.5). The wavelengths that we see are determined by the physics of the eye and the chemistry of the photoreceptor pigments in the eye. Electromagnetic energy with wavelengths below the visible spectrum includes ultraviolet rays, X-rays, and gamma rays; electromagnetic energy with wavelengths above the visible spectrum includes infrared rays, microwaves, and radio waves.

Our sensitivity to spectral radiation is not constant across the electromagnetic spectrum. Instead, we sample it through a pair of filters. The first filter is provided by the spectral sensitivity of the rods, which have a maximum sensitivity around 500 nm (498 in Fig. 4.5). The second is provided by the pooled responses of the three types of cones (the leftmost and two rightmost curves in Fig. 4.5) and has maximum sensitivity around 555 nm. Below 380 nm (infrared) and above 700 nm (ultraviolet) we are effectively blind to electromagnetic waves. Therefore, there is a wide range of information that meets our eyes which falls outside our window of visibility.

Although we sense infrared energy as heat, other wavelengths outside the visible spectrum are imperceptible without the aid of devices such as radios or infrared goggles. Ultraviolet energy is destructive to living tissue, so it is filtered out by the yellow pigment in the lens of the eye. People who have had their lenses removed because of cataracts, however, can see ultraviolet energy as light. Rather than experiencing it as a new color sensation, which would require another type of cone, they see it as the same color that people with normal vision would see violet.

Blind spot   **1 2 3 4 5 6 7 8 9 a b c**

Fig. 4.6 Cover your right eye and focus directly on a digit with your left eye. Then move the page away from your head. At about a foot the word blind spot will disappear. The location of the blind spot and screen sizes will vary, so you may have to try different focus points

When you switch from looking at colored objects in bright light to looking at the same objects in dim light, all other things being equal you will notice that one object now seems brighter than the other. If you look at a blue flower and a red flower, for example, first in daylight and then at dusk, you will notice that under low illumination both appear faded, but the blue seems brighter than the red. Likewise, a piece of green paper and a piece of red paper which are matched for brightness in good light will not be matched in dim light. This effect is called the Purkinje Shift, and is based on the fact that long wavelength colors such as red appear duller under low illumination than shorter wavelength colors (such as blue). The effect occurs because of the shift from high illumination vision (cones) to low illumination vision (rods) under different light conditions. The rods are relatively more sensitive to light in the blue region than the cones, hence the apparent greater brightness of the blue flower in dim light. These relationships can be seen in spectral sensitivity curves that illustrate that maximum sensitivity goes from red to blue green (i.e., to shorter wavelengths) when we shift from bright to dim light and from the cones to the rods.

One final point about rods and cones: there is one part of the retina where there are no receptors (rods or cones) present. This is where the optic nerve leaves the retina. When the image of an object falls on this *blind spot*, nothing is seen. You can use Fig. 4.6 to help you find your own blind spot. The blind spot is in a different place for each eye.

4.3.5 Implications for System Design

If your system is going to be deployed in an area where users experience large step changes in lighting conditions on entering that area (from light to dark or vice versa) you will need to take account of the fact that their eyes will need time to adapt to the new lighting conditions. You could consider having gradual changes in lighting levels as they enter the area, for example.

People will often use a relatively large dialog box positioned in the center of the screen to get the user's attention for something that is important. This dialog box is usually large enough to project an image onto the retina that fills a significant portion of the fovea. It would be equally possible to capture the user's attention by make an item on the display screen move, or make it flash or blink, as long as

the item is located in a position that projects onto the periphery of the retina. In this way the movement or flashing gets detected by the rods in the eye. There is a negative implication here too: if your user is carrying out an important task that requires high levels of concentration using the display screen, you should try to avoid having items on the screen move or flash, particularly if they would project onto the periphery of the user's eye.

If you are designing a system that has to be operated in light sensitive conditions, such as a photography dark room or a radar operations room, you need to consider how you can help the users see so that they can carry out their tasks in the dimly lit conditions. Usually in these situations, red light is used to illuminate the rooms, albeit at quite low levels. The rods are relatively less sensitive to light at the red end of the visible spectrum, which means that the rods start dark adapting even though there is still some light available.

In most cases people using technology work at arm's length from their display screen. This means that the blind spot is usually not a problem, because the brain processes the images from both eyes together to perceive what is on the screen. If you have a system which requires the user to be much closer to the screen, and they are not detecting some items that are important, you should consider investigating whether those items are in fact located in their blind spot.

4.4 Low Level Visual Perception

Here we consider the low level details of vision. These range from how light is detected, through various aspects associated with color, to flicker and pop-out effects. Several of these aspects of low-level vision have implications for system design.

4.4.1 Vision and the Measurement of Light

There are two important ways in which light gets from an object to the eye: incident light (light falling on an object) and reflected light (light reflected from an object). Incident light is referred to as *illuminance* whereas reflected light is termed *luminance*. White surfaces typically have reflectances of 80% and black surfaces around 10%.

Luminance is measured in candelas per square meter (cd/m^2). As the luminance of an object increases, so does the eye's visual acuity or ability to discern small details. The pupil's diameter decreases and therefore increases the depth of focus in the same way as a standard camera lens when the aperture is adjusted. An increase in luminance of an object or display will also make the eye more sensitive to flicker.

Contrast describes the relationship between light emitted from an object and light emitted from the surrounding background. It is defined as the difference

Fig. 4.7 A Hermann grid on the *far left*. In the *center* a similar effect in a file system. Adding shading and color removes the effect on the *right*

between the luminance of the object and its background divided by the luminance of the background, as shown in Eq. (4.1).

$$(\text{Object Luminance} - \text{Background luminance})/\text{Background Luminance} \quad (4.1)$$

The contrast will be positive if the object is emitting more light than the background and negative if the background is emitting more light than the object. Objects can therefore be described as having positive or negative contrast.

Brightness is a subjective response to light. There is no real means of measuring absolute levels of brightness but, in general, a high luminance from an object implies a high brightness. It is possible to experience odd effects at high-to-low brightness boundaries, as shown in the left part of Fig. 4.7, which is called a Hermann grid. Designers should be wary of creating effects like the Hermann grid (in the center of Fig. 4.7) because they can be distracting. Adding color or shading can remove this effect. Indeed, the design of folders on most modern desktops avoids this by increasing the spacing of the icons both vertically and horizontally, but this choice costs display space.

Related to the Hermann grid are the concepts of *figure* and *ground*: figure refers to the objects that are to be attended to; ground refers to the background objects. In Fig. 4.7 the boxes and files are the figure and the white background is the ground. When the objects are placed too close together, however, the perceived gray fill that results can become more prominent and appear to be part of the figure. It is important to consider keeping objects that are figure prominent, and maintaining a useful ground.

The interpretation of Fig. 4.8 is based on figure and ground. If the white is perceived as the figure, it appears to be a vase or a goblet; if the black is the figure, it appears to be two heads.

The objects that you want the users to see or distinguish need to be appropriately sized. In good viewing conditions a minimal perceptible visual angle of about 15 min of arc should be maintained and in poor viewing conditions this should be increased to 21 min. These correspond to a 4.3-mm object and a 6.1-mm object, respectively, viewed from 1 m.

The full *field of view* for a stationary forward looking eye covers about 208° horizontally (although it is blocked by the head and the nose at certain points), and about 120° vertically. This only refers to light falling on the eye, and does not

Fig. 4.8 A Rubin vase that
can be interpreted as a vase or
two heads facing each other

necessarily mean that something will be seen when the signals are further processed. The field of view is an important factor in determining the size of a particular display screen or the layout of displays and control equipment.

4.4.2 Color Vision

Color is the result of perception, and is not an intrinsic part of an object. The corollary of this is that, under different lighting or contrast conditions, the apparent color of an object will change. There are some terms that you should be aware of when thinking about color vision.

Lightness is a measure on the black–white dimension of how close to white the color is (100% is white, or bright). The amount of color that there is in the light is described by its *saturation*, which refers to the purity of the sensation as opposed to grayness: the higher the saturation, the higher the purity of color. *Hue* is primarily dependent on wavelength; it is closest to the way that the term *color* is used in everyday language.

If the wavelength of visible light is varied over its range (400–700 nm), with constant luminance and saturation, a person with normal color vision can distinguish about 1,200 differences in color. If luminance and saturation are also varied, approximately 8,000 differences in color can be distinguished. When the viewing is carried out in isolation by a person with normal color vision, however, only 8–10

different colors can be identified accurately without training. If you need more than 8–10 distinct colors you may need to use texture or text to provide features to assist the user in discriminating between them.

People's sensitivity to color is not uniform across their field of view. The eye is not sensitive to color at the periphery of vision. Accurate discrimination of color is only possible to around 60° from the straight ahead position (with the head and the eyes stationary) and the limit of color awareness (as opposed to discrimination) is approximately 90° from the straight ahead position. The eye is best suited to the perception of yellow-green light, and color is only well perceived in foveal (central) vision. It is least sensitive to red, green, and yellow light at the periphery of color vision where it is most sensitive to blue light. This variation in sensitivity arises from the way that the rods and cones are distributed in the fovea.

Perceptions of a particular color are affected by prolonged exposure to other colors—this is because different cones are responsive to different dimensions of color (e.g., red–green or yellow–blue). Looking at red light, for example, causes the red cones to become adapted, so the red light reduces in salience. This is often seen in color after effects, or afterimages. There are also several visual illusions you may have seen where you first stare at a picture or a display until the colors apparently disappear, and then when you look at another picture or a blank piece of paper the complementary colors appear.

Color constancy refers to the situation in which we attempt to perceive colors as being the same even when they are different. Our clothes do not change color when we go indoors, for example. The wavelengths hitting the retina may have changed, however, and the reflected light will be different, although we will still perceive the colors to be the same. Vegetables and meat in supermarkets are one of the most compelling examples of constancy not working—stores tend to be lit so that objects give off particular wavelengths—when you get home with different lighting the colors are substantially duller.

4.4.3 Color Blindness

It is important to be aware of color blindness (or, more strictly color vision deficiency) because around 7% of Western men and 0.5% of Western women are red–green color deficient. In other words they are bichromats, because they can only distinguish two primary colors (typically they cannot tell red from green). Other cues, such as brightness, can be used to help distinguish red from green.

Most of us can discriminate between all three primary colors (red, green, and blue), and are known as trichromats because we have three types of cones for seeing color. Quadchromats appear to have better color discrimination under different lighting conditions, but see the same colors. Many animals are monochromats: they cannot distinguish any colors because they have only one set of cones (or just rods). They hence perceive the world in monochrome.

4.4.4 Color Systems

There are many theories of how we perceive color (Sekuler and Blake 2005). There are a lot of technical details and tools available which allow us to make quite strong predictions about color perception. For our purposes, the most important thing to remember is the high level distinction between the two ways that we perceive color: additive and subtractive.

Computer displays use projected light (for the most part) and printed paper uses reflected light. Projected light uses a different set of primary colors, the additive colors of red, green, and blue (often referred to as RGB), while reflected light use the subtractive colors: cyan, magenta, and yellow (with black as the *key* color, which is why they are often referred to as CMYK). The additive colors start from black, no light, and then add colors to get other colors, and then, when all are added, white is produced. The subtractive colors remove colors from white, effectively reducing the amount of light that is reflected, and ending up at black. Thus, with current technology it is hard to get a pure black color on a display screen (at least some of the lights tend to stay on), although it is possible to get closer to pure black using subtractive colors.

The colors from printed materials cannot completely match the colors from a screen because of the way the two different color systems work. There are some tools available that can help you make the two match more closely, however, but in the end the choice of how closely you make them match may be down to personal preference.

4.4.5 Flicker

People are sensitive to flickering lights. The flicker can be made imperceptible, however, by using rates that exceed the flicker fusion rate. In the early days of the movies, cinema films ran at a frame rate below the flicker fusion rate, which is why any movements appear jerky and you can still detect the flicker (which explains why they were often referred to as the flicks). In modern cinema, the flicker problems is overcome by typically showing films at 24 frames per second, with the image being presented twice per frame on two-blade shutter projectors, and three times per frame on three-blade shutter projectors.

It should be noted, however, that lights which flicker at a rate of 7–10 cycles per second[1] can trigger epileptic fits in some people. Series 500 of the ISO 9241 standard on ergonomics of human–computer interaction is a guideline that addresses this issue for display screen equipment, and you sometimes hear warnings about flickering lights and images before they appear during a television

---

[1] You will also see this measure in Hertz: 1 Hz is 1 cycle per second.

C C C C C C C C C   C C C C C C C C C C

C C C C C C C C C C   C C O C C C C C C C

C C C C C T C C C C   C C C C C C C C C C

C C C C C C C C C C   C C C C C C C C C C

C T C C C C C C C C   C C C C C C O C C C

C C C C C C C C C C   C C C C C C C C C C

Fig. 4.9 The *left half* of the figure shows a bunch of Cs and Ts. The *right half* shows a bunch of Cs and Os. You should be able to see that the Ts pop-out from the Cs whereas the Os do not

program. The effect of flicker in system design is decreasing as the refresh rate of most displays are so fast that flicker is becoming less of a problem.

4.4.6 Pop-Out Effects

One of the most useful applications of vision to interface design is to take advantage of how the eye searches. One of the most useful effects is that certain stimuli 'pop out' from other stimuli. Figure 4.9 shows that with a bunch of Cs, Ts pop-out, and Os do not pop-out. Similar effects can be found for color, and a few other features (Treisman and Gelade 1980).

Figure 4.10 shows an application of this and a few other effects. First, you have some object recognition that occurs—a car and building and perhaps some signs are recognized. Then you have some automatic processes occur, you are expert at reading so words appear as well, and these pop-out to a certain extent. You also have the ability to read words, and these appear. Expertise and previous experience also counts. If you have not heard of French Connection—UK (FCUK), your automatic processes may read it as something else. This result is not a true pop-out effect, but it is a related phenomenon, the word completion effect.

Providing multiple encoding of search items also helps in a visual search (Garner 1974). In Fig. 4.11 targets are encoded in several ways using multiple features. Highlighting visual targets using just two discriminating features often helps users find them, particularly if the encoding is not taught to users but is just inherent in the interface.

Some feature differences will pop-out to users (Treisman and Gelade 1980). Differences in color, when there is not much color, and straight line segments in a set of smooth curves will both pop-out. Partially open circles will not pop-out when mixed with circles and slight changes in size will not pop-out. Figure 4.11

Fig. 4.10 Sign for French Connection-UK, taken in Bath, England. Note the effects of automatic processing, the word completion effect, on multiple sets of letters in this picture

Fig. 4.11 A picture showing a bunch of Cs and a Z and an O with the Z and O in bold, highlighted, slightly larger to provide three features that can pop-out. The Z has more features different so it will pop-out more easily. The O without the shading would be harder to find

has the targets modified with some differences that will pop-out and some differences that will not pop-out.

Pop-out effects can be used to highlight spelling mistakes (as shown in Fig. 4.12), to highlight a target file within a folder, and to highlight target words on web pages during a search.
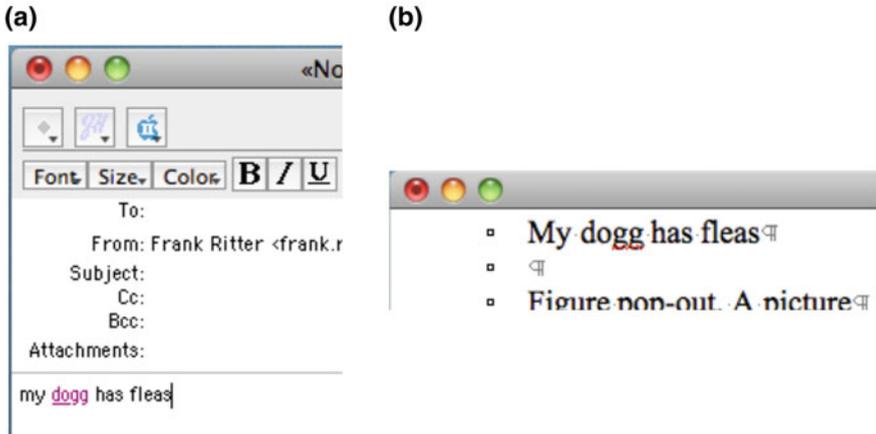
Fig. 4.12 Example uses of visual encoding to highlight spelling mistakes. On the *left*, a mail system underlines and makes the misspelled word appear in *red*; on the *right*, Microsoft Word puts a *wavy red line* under the word

Fig. 4.13 The use of color and luminance on aircraft displays facilitates the finding of particular information (this will be less apparent in b/w versions of this figure!)

4.4.7 Implications for System Design

It has long been a rule of thumb that you should use color sparingly on display screens. In other words, you just use color to emphasize the things that are important. If your system has displays that rely on widespread use of color, and there is meaning associated with the different colors, you should check to make sure that the exposure to the dominant color on the screen is not adversely affecting the perception of some of the other (possibly more important) colors.

When you do use color to encode items on a display, you need to make sure that the users can perceive the colors that you are using. Figure 4.13, for example, shows how color can be used to help users find information. The artificial horizon display above the throttle levers has blue sky and brown ground. The center displays have green and yellow and orange text to help separate information. If you are designing safety critical systems or interfaces used by a wide range of users, you should consider using redundant information to help people with red-green color vision deficiency distinguish between these two colors on the screen.

If you are developing a system which must produce color print-outs, and the colors and the differences between colors are important, you will have to take into account the fact that the colors on the screen and on the print-out will appear somewhat different. You will therefore need to find a way to make the two representations match as closely as possible, which may involve using third party software. This process may be further complicated by the fact that different makes of color printer can produce print-outs of the same screen image that look different.

Nowadays the flicker rate of most display screens is high enough for the flicker to be imperceptible. There is some evidence, however, that the flicker rate can affect how people read, because it changes the size of the saccadic eye movements that take place during reading (e.g., see Kennedy and Baccino 1995). If you are developing a system which requires people to read large amounts of text from the display, you may want to make sure that you choose display screen hardware that has a flicker rate that minimizes the amount of interference on reading.

If you are designing a system where you need the user to focus on one (or more) particular items on a densely populated screen, you should consider whether you can make those items pop-out from the display by appropriately changing some of the items' features. There is a potential downside of working with densely populated displays too, in that there may be some items which pop-out of the display when you do not want them to. In these cases you would have to consider how you could make those items appear more similar to the surrounding items.

4.5 Higher Level Visual Perception

In addition to how low level aspects of the eye influence perception, there are several higher level aspects of vision that influence cognition. Several of the

results here have different interpretations, as to whether they are the result of vision or the result of cognition. A comprehensive theory of the user would incorporate this interaction between vision and cognition. Being aware of the higher level perceptual issues can help you design better interfaces.

4.5.1 Movement and Spatial Perception

Movement can be detected either by moving ourselves (even though the image falls on the same part of the retina) or by staying still whilst the image moves across the retina. The processing of the stimuli related to body movements and visual stimuli are combined because we can track a still object with our eyes while moving our bodies and yet be sure that it is still.

   Above a certain speed of movement the eye can spontaneously track a moving object. Ask someone to slowly wave a small light (about as bright as a lit match) around in a darkened room and follow it with your eyes. The movement of the light will be seen even though no image is moving across the retina (because your eyes will keep it constantly on the retina). The periphery of the retina is the area most sensitive to movement, but it is very difficult to identify an object at the extreme periphery of the field of view. The detection of a moving object in the periphery of the eye is what usually initiates the movement of the eye in pursuit of that object so that it can be brought into focus on the fovea. Movement may also be perceived as afterimages when both the eye and the retinal image are stationary (this effect is due to adaptation of the motion detectors in the eye). Thus, the eye and the brain combine their results to get this apparent motion.

4.5.2 Depth Cues

Spatial perception is determined from the muscular activity of the two eyes and discrepancies between the two images that are formed. When we want to display a 3D image on a screen, we have to represent the 3D image using just two dimensions, although better 3D displays continue to appear. We can simulate depth perception using the discrepancy between the two images, as well as the perceptual depth cues listed in Table 4.3.

   In the real world, motion parallax may be one of the most important cues that enable us to perceive distance and depth. It occurs when we move our heads from side to side and we see objects displaced at different rates. Objects that are further away appear to move more slowly than objects that are closer. In screen design, the trick is to move the viewpoint of the "camera" so that the image on the screen moves according to the principles of motion parallax. This is not used in most non-game interfaces, although virtual reality systems provide it, and video games provide it through motion of objects.

Table 4.3 Perceptual depth cues

Size	The larger of two otherwise identical objects appears to be closer than the smaller one
Interposition	If one object partially occludes a second object then the blocked object is perceived to be behind and beyond the blocking object
Contrast, clarity and brightness	Sharper and more distinct objects appear to be nearer, and duller objects appear to be farther away
Shadow	Shadows cast by an object provide some cues about the relative position of objects
Texture	As the apparent distance increases, the texture of a detailed surface becomes less grainy
Motion parallax	When moving one's head from side to side the objects one sees are displaced at different rates
Stereoscopic depth	Two images of the same object from slightly different angles are presented separately to each eye. Perceived depth is induced through the fusion of the two images. This is often used in virtual reality

Fig. 4.14 The time to count objects as the number of objects varies from one to eight. Counts from one to three are substantially faster per item than for five and greater (reprinted with permission from Peterson and Simon 2000)
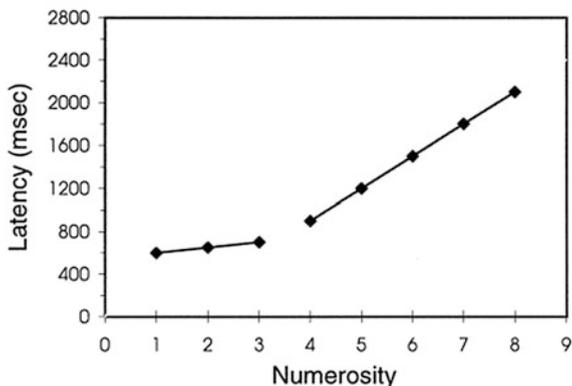
4.5.3 Subitizing

Figure 4.14 shows that the curve for counting objects presented visually in a psychology experiment (not that appear one at a time like train cars or a haptic situation like counting coins in your pocket) has a bend on it at about between 3 and 4. Up to three objects, you recognize the number effectively, with about 50 ms difference per object, above four and certainly at five, you have to count the objects, so the time to respond increases by about 250–300 ms/object. The first part of the curve is called subitizing, and is thought to be an effect of the architecture of the perceptual system.

Given the effect of subitizing, if you are passing sets of information to a user, or giving them things to count, understand that smaller numbers are much faster to recognize and count than larger numbers. In a related effect, if you are counting

Table 4.4 The Gestalt principles of visual grouping

Proximity	Elements that are close together appear as groups rather than as a random cluster of elements
Similarity	Elements with the same shape or color are seen as belonging together
Common fate	Elements which appear to move together are grouped together
Good continuation, continuity	Elements that can be grouped into lines or shapes will be
Closure	Missing parts of the figure are filled into complete it, so that it appears as a whole
Symmetry	Regions bounded by symmetrical borders tend to be perceived as coherent figures
Figure-ground	The geometric organization that is perceived is the one with the best, simplest, and most stable shape. For example, four dots arranged as if they were at the corners of a square will be perceived as a square rather than a triangle plus an extra dot

based on a fixed number of objects you can track, then you cannot expect users to follow more than three or four objects on the screen.

4.5.4 Gestalt Principles of Grouping

The *Gestalt principles of visual grouping* (listed in Table 4.4) can be used to explain how groups of objects are interpreted. The principles were developed as a rebellion against the simplistic notion that perception could be structurally analyzed into its component parts, and that complex ideas were the result of associating together simpler ones. This simplistic view dominated psychology in the late nineteenth and early twentieth centuries. The Gestalt principles allow that the whole can be more than just the sum of the parts.

Gestaltists focused on the fact that there are important aspects of form and structure. We see the world as composed of discrete objects of various sizes that are seen against a background comprised of textured surfaces. The spatial and temporal relationships between elements are as important as the absolute size, location, or nature of the elements themselves, and a sensation-based account of perception fails to capture this. Some of the principles of visual grouping are illustrated in Fig. 4.15.

### 4.5.5 Other Theories of High Level Visual Perception

There are several other views of visual perception apart from the psycho-physiological one we have described. These views tend to deal more with perception-as-phenomenological-experience or perception-in-information-processing rather than
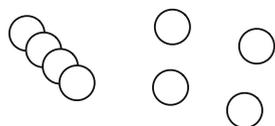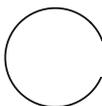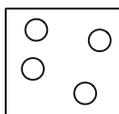
**Proximity groups
the left hand dots**

**Similarity makes the left hand
dots belong together more**
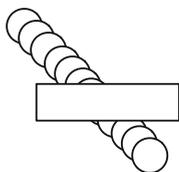
**Good continuation makes the line of
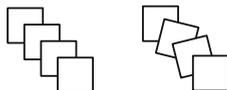dots appear to continue through the rectangle**

Closure makes the circle appear whole

**Surrounded by a box,
the dots look more like a group**

**Symmetry and equal size make the dots
look more like a group (and helps the
illustrations and captions group in this figure)**

**Orientation makes the dots on
the left look more like a group**

Fig. 4.15 Several illustrations of the Gestalt principles of visual grouping

perception-as-registering-and-coding-sensation. They include the Constructivist, the Ecological, and the Active Vision approaches.

In the Constructivist approach, seeing is regarded as an active process in which our view of the world is constructed from a combination of information in the environment and previously stored knowledge. So, what we get from our eyes is not a 2D visual representation of the world like a photograph. Instead, the visual system constructs a model of the world by transforming, enhancing, distorting, seeking, and discarding information. In doing so, the visual system provides us with a much more constant view of the world than if we were simply to "see" the images produced on our retinas. So, when we move about (e.g., walk down a street), buildings appear stationary and people appear to be approximately the same size and shape—despite the fact that their actual relative images on the retina

may be quite different. In the same way, our ability to perceive objects displayed on computer screens, for example, whether they are text or graphics, 2D or 3D representations, is a result of our prior knowledge or our expectations as to what should appear as well as what lands on our retinas.

In the Ecological approach (e.g., Gibson 1979), the process of seeing is greatly influenced by what uses (affordances) the object perceptions suggest. This approach, which is also called direct perception, stems from work with aircraft pilots. It takes as its starting point not a retinal image that is passively sensed, but the ambient optical array that the observer samples. Perception and actions are seen as tightly interlocking and mutually constraining.

In the Active Vision approach, Findlay and his colleagues (e.g., Findlay and Gilchrist 2003) build on the Ecological approach, but emphasize the role of cognition in controlling vision. This view opposes the idea of just taking scenes and analyzing them, and instead incorporates the notion of interacting with scenes. In interacting, the eye (and brain) chooses where to look next based on what has been focused on before and what was seen in those locations.

4.5.6 Implications for System Design

If you are designing a system where the three-dimensional positioning of items on the display is important, you will need to decide which perceptual depth cues you should use. The final choice is likely to depend on the context in which the display screen will be used, as different lighting conditions can affect the way that the displayed image is perceived, so if you only use narrow and quite pale shadows, for example, these may not be very obvious in brightly lit conditions.

For critical systems, if the user has to undertake a task which involves responding to the number of target items that appear on the display screen, you will need to consider how to display those target items. If the number of items is very low, the users will be able to react very quickly. If the number is very large, however, you may need to think about organizing them in some way (using the Gestalt principles, perhaps), or even splitting them across displays to optimize the response time.

At present, most display screens offer one level of resolution, such that all items on the display are presented with that resolution irrespective of their importance. If you are designing a system which requires an increased level of detail for items that are currently being used (perceived) then you may want to consider the possibility of using gaze-contingent multi-resolution displays (Reingold et al. 2003). These displays maximize the resolution at the part of the screen where the user is currently looking. This approach can provide a much more detailed image for a given bandwidth because more pixels (and information) are allocated to what the user is currently looking at, rather than effectively wasting pixels by providing an unnecessary level of detail for items that the user is not currently looking at.

When people are designing displays which involve items that need to be grouped together they often include them within some form of container (usually a box). The Gestalt principles suggest, however, that careful placement of the items will also be sufficient to determine how they are perceived by the user. The laws also suggest that careful attention needs to be paid to how you lay out items on a display, because you can, unwittingly, end up with unrelated items being perceived as being related simply because of their relative placement.

4.6 The Auditory System

For normally sighted and hearing people, hearing is the most important sense after vision in any interaction. Most people can hear sound in the frequency range 20 Hz up to 20,000 Hz, but both the upper and lower frequency limits tend to deteriorate with age and health. Hearing is more sensitive within the range 1,000–4,000 Hz, which in musical terms corresponds approximately to the top two octaves of the piano keyboard, and represents much of the range of the human voice.

Thus, the stimulus for audition is any vibration that will set the ossicles (small bones) of the ear in motion between about 20 and 20,000 Hz. Ordinarily, this means vibrations of the air but vibrations transmitted through other bones (particularly the skull) also contribute to auditory sensation. (Having a tooth extracted or drilled will almost convince you that the jaw was designed to transmit vibrations to the ear in the most efficient manner possible!) There are now headsets available that use direct bone conduction as a way to transmit sound in noisy environments.

4.6.1 Theoretical Description of Sound

It is convenient to consider the stimulus for sound to be made up of successive compressions and rarefactions (expansions) of air that follow a waveform over time. An example is shown in Fig. 4.16.

Waveforms like that in Fig. 4.16 can be summarized as being made up of sine waves (they look a lot like the first part of the waveform in Fig. 4.16 but are smoother and follow the sine function used in trigonometry). There are at least two reasons for using the sine function. The first is that they are easy to create; a pure tone produced by an electronic oscillator or a tuning fork follows a sine wave. The second and more important reason is that theoretically a wave of any shape can be analyzed into component sine waves. This is known as Fourier analysis. Figure 4.17 provides a simple example. Work with sine waves thus provides a standard for comparison across different types of sounds.
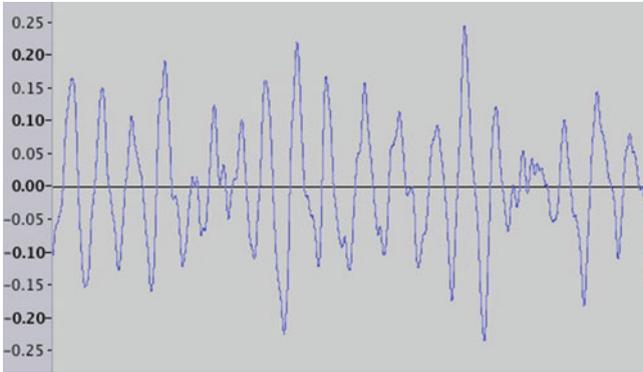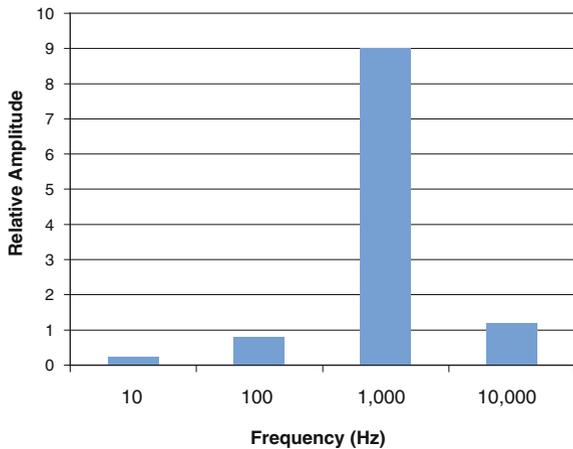
Fig. 4.16 An example waveform from audacity

Fig. 4.17 Example, simplistic, Fourier analysis that might match the waveform in Fig. 4.16

Waveforms with a single sine wave sound purer. Waveforms with different combinations of frequencies sound different. How many of these subwaves there are determine whether you are hearing a piano or an organ, or Frank or Gordon speaking.

Loudness of the sensation is largely dependent on the amplitude of the wave. However, the ear is not equally sensitive to all frequencies, so sounds at different pitches will not have the same loudness. The pitch of the tone depends primarily on the frequency of the sine wave, but not completely. Pitch is also dependent on amplitude. The apparent pitch of high frequency tones will increase with increasing amplitude but the apparent pitch of low tones decreases with increasing intensity. The loudness of a tone will also depend on the phase relationships of the component frequencies of the stimulus (that is, do they all start at once or do they start after each other but offset in time). Timbre is a quality that depends on the

purity of the tone; is it made up of one single sine wave frequency or a broad mixture of frequencies? A tuning fork has a relatively pure tone and therefore little timbre. On the other hand, a piano or other musical instrument has timbre because of the other frequencies present in its sounds. Different timbres are often assigned different meanings by users, which may be important for your design.

Low tones of equal loudness appear to occupy more space and thus are said to have more volume than high tones. On the other hand, high tones have a greater density than low tones of equal loudness. The volume and density of tones are each a joint function of intensity and frequency of the tones. However, they seem to be as real as pitch and loudness which have simpler bases. In other words, listeners have no difficulty making reliable judgments of volume or density of tones that differ in frequency and intensity.

Hearing can be likened to carrying out a type of Fourier analysis of the auditory stimulus, separating a complex wave into its sine wave components. There are some situations where this analogy breaks down, such as when two stimuli of approximately equal intensity and frequency are simultaneously presented to the ear. Instead of hearing both tones, as a linear Fourier analysis would allow, a single tone is heard which varies in loudness in a periodic manner. You may have heard this when two people sing together or two instruments are played together. The effect can be pleasant or unpleasant depending on the frequency of the beats.

The basis of beats is the following. If you have a pure tone of 256 Hz and another of 257 Hz, each one would produce a steady pitch that would be difficult to distinguish from the other. When the two are played together the compressions and rarefactions (expansions) of the air produced by the two tones will at some point be in phase (synchronized) and the two tones will add together. However, because the frequency of one is slightly greater than the other, they will get out of phase after a while and their effects will cancel each other out. As this process repeats, they will go in and out of phase as many times per second as the difference between the tones in cycles per second. In this example, it would be once per second and so you will hear one beat per second. This provides a very accurate way of measuring the difference between two tones, far better than the ear could discriminate if the two tones were presented separately. This fact is used to good effect by piano tuners. They tune one note until it no longer beats with the standard tuning fork. Then the other notes are tuned until their harmonics do not beat with the first note.

4.6.2 Measuring Sound

Sound intensity is normally measured using the *deciBel* scale. This is a relative logarithmic scale where 10 decibels (dB) = 1 log unit ratio of energy, or a Bel (name after Alexander Graham Bell). To give some practical examples, the threshold of hearing is 0 dB, a whisper registers 20 dB, and normal conversation registers between 50 and 70 dB.
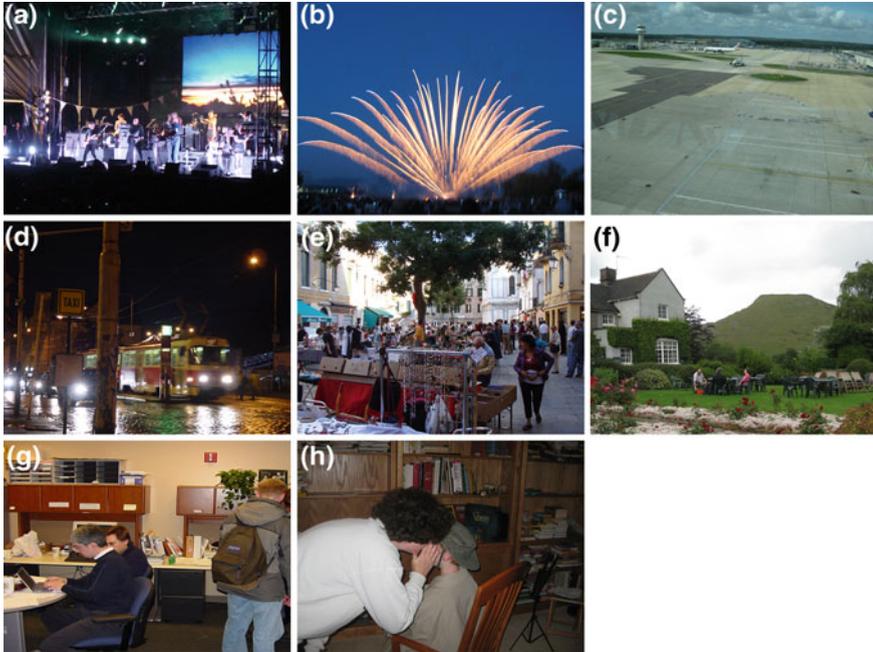
Fig. 4.18 Example sound levels (in decibels). **a** [140] Rock band (amplified) at close range; **b** [120] Loud thunder or fireworks; **c** [100] Jet plane at 500 ft; **d** [100] Subway train at 20 ft; [90] (not shown) Potential for permanent hearing loss; **e** [80] Busy street corner; **f** [60] Normal conversation; **g** [40] Typical room; **h** [20] Whisper; [0] (not shown) Threshold of hearing

The intensities of various common sounds are shown in Fig. 4.18. The sensitivity to both frequency and loudness varies from person to person. Generally, though, the ear is insensitive to frequency changes below about 20 dB (i.e., below a whisper). Once levels get above 90 dB (as shown by the line in the table), prolonged exposure can lead to permanent hearing loss. This level is often surpassed by some industrial jobs, and by the iPods and MP3 players of those people who have the volume control turned up excessively high.

This scale does not describe perceived loudness well. Increasing an auditory stimulus by equal ratios does not produce equal increments in sensation. It is obvious that the difference in loudness between a whisper and a normal conversation is less than the difference between a normal conversation and a subway train. What is said, the context, and background noise also influence perceived loudness. However, the ratio of the energy in a whisper to that in a conversation is about the same as the ratio in conversation to that of the noise of a subway train. This explains why a 100-W stereo is much louder than a 1-W pocket radio, and a 200-W stereo is not much louder than a 100-W stereo. The corollary of this is that when measuring loudness, asking people to directly estimate apparent intensity of

loudness (known as magnitude estimation) is often the best way to quantify sounds, at least for interface design.

4.6.3 Localizing Sound

People are generally poor at using spatial cues to successfully localize sounds (Catchpole et al. 2004; Kubovy and van Valkenburg 2001). The idea that localization is based on inter-aural time differences at low frequencies and inter-aural intensity differences at high frequencies is called the 'duplex' theory and dates back to Lord Rayleigh (1907), a pioneer in perception. This does not hold for complex sounds.

We can identify the location of a sound from the time taken for waves to reach the ears, coupled with information from head and shoulder movements. Sound reaching the far ear will be delayed in time and will be less intense relative to that reaching the nearer ear. Thus, there are two possible cues as to the location of the sound source. Owing to the physical nature of the sounds, these cues are not equally effective at all frequencies.

Low frequency sounds have a wave length that is long compared with the size of the head, and this "bends" the sound around the head very well. This process is known as diffraction, and the result is that little or no shadow is cast by the head. On the other hand, at high frequencies where the wavelength is short compared to the dimension of the head, little diffraction occurs. A "shadow" almost like that produced by a beam of light occurs.

Inter-aural (between-ear) differences in intensity are negligible at low frequencies, but may be as large as 20 dB at high frequencies. This is easily illustrated by placing a small transistor radio close to one ear. If that ear is then blocked with a finger, only the sound bending around the head and entering the other ear will be heard. The sound will be much less "tinny" because high frequencies will have been attenuated more than low; the head effectively acts like a low pass filter (allowing only low frequency sounds). Inter-aural intensity differences are thus more important at high frequencies than at low ones.

If a tone is delayed at one ear relative to the other, there will be phase differences between the two ears (the peaks of the waves will arrive at different times). If nerve impulses occur at a particular phase of the stimulation waveform, the relative timing of the nerve impulses at the two ears will be related to the location of the sound source. This is used to locate "wide" sounds. However, for sounds whose wavelengths are comparable with, or less than, the distance between the two ears there will be ambiguity. The maximum path difference between the two ears is about 23 cm, which corresponds to a time delay of about 690 μs. Ambiguities occur when the half wavelength of the sound is about 23 cm, i.e., when the frequency of the sound is about 750 Hz. A sinusoid of this frequency lying to one side of the head produces waveforms at the two ears that are in opposite phase (phase difference

between the two ears of 180°). From the observer's point of view, the location of the sound source is now ambiguous, because the waveform at the right ear might be either a half-cycle behind that at the left ear or a half-cycle ahead. Head movements or movement of the sound source may resolve this ambiguity, so that there is no abrupt upper limit in our ability to use phase differences between the two ears. However, when the wavelength of the sound is less than the path difference between the two ears, the ambiguities increase; the same phase difference could be produced by a number of different source locations.

There are two different mechanisms for sound localization: one operates best at high frequencies and the other at low frequencies. For middle frequencies neither mechanism operates efficiently, and errors are at a maximum. Stevens and Newman (1936) investigated localization of single bursts with smooth onsets and offsets for observers on the roof of a building so that reflection was minimized. The listeners had to report the direction of the source in the horizontal plane, to the nearest 15°. Although left–right confusions were rare, low frequency sounds in front were often indistinguishable from their mirror location behind. If these front–back confusions were discounted, then the error rate was low at very low and very high frequencies and showed a maximum for mid-range frequencies (around 3,000 Hz). Intensity differences are more important at high frequencies, and phase differences provide usable cues for frequencies below about 1,500 Hz.

4.6.4 Discriminating Sounds

Our abilities in discriminating sound depend upon whether we mean absolute discrimination or relative discrimination (this applies to vision too). Absolute discrimination is quite poor (e.g., systems should not rely on remembering a tone or sound), but relative discrimination is very good. With sounds we can remember no more than five to seven items for absolute discrimination unless we can attach meaning to them, such as pitch labels. Also, as we vary more of the dimensions of the stimulus (increasing its complexity), so we increase our ability to discriminate (up to 150 sounds—varying in frequency, rhythm, location, duration, volume, etc.).

4.6.5 Implications for System Design

There are two ways in which sounds are used in current systems. The first is to provide voice output. The second is to provide audible alerts, such as telephone ring tones, and audible alarms.

Voice outputs generally require more processing than plain sounds. They can convey much more information, however, and they are particularly important for people with impaired vision. Blind users, for example, use voice output with screen readers so they can process the text shown on a display screen.

In general it is said that automated female voice output is easier to understand. If you are designing a system that requires voice output, the rule of thumb is usually to prefer a clear, slightly high, female voice if you can only choose one. If you have a system that produces multiple voices in the same task context, however, you will need to think more deeply about your choices. First, and foremost, however, you will need to choose voices that are easily discernible from each other. If your system has a sophisticated speaker system, you may also be able to separate the voices spatially.

You will also need to take account of the geographical context in which your system will be used. If your system is being deployed in the UK, for example, users may prefer the system to use a voice that has a British accent. Personal preferences can play a role too, so you may want to allow your user to select which voice output should be used. Indeed, many motor vehicle satellite navigation (satnav) systems now allow you to choose the voice that will be used to offer you verbal directions on how to reach your travel destination.

Like sight, hearing generally diminishes with age. If you are designing a system for a population that will include older users, you will need to take the possible reduced hearing levels into account when selecting voices and sounds to use as output. Some mobile phones by default include ring tones that can compensate for hearing loss in older people.

Audible alarms are supposed to be designed to alert the user to abnormal or undesirable situations. In many cases the audible alert is used in addition to a visual alert. If you want to quickly attract your user's attention in an emergency situation, the easiest way is to use square-ended, sudden onset waveforms such as klaxons or bells to provide audible alarms. The problem is that they evoke a startle reaction, and if the sound is at high volume it can even cause panic. The design of audible alerts requires great skill because you want to make sure that your users process the meaning of the alert rather than focusing on trying to stop the loud, incessant noise. The alarm sound needs to be distinctive so that users can recognize it and recall its meaning. If you are trying to convey urgency, you should avoid high intensity sounds, and instead consider the speed of the alarm sound: a tri-tone alarm, for example, which doubles in speed, will suddenly sound very urgent.

4.7 Motivation

4.7.1 Introduction

Now that you know some of the basics about *how* humans behave, it is also important to be aware of *why* they behave in particular ways. Generally there are two reasons why people act in a particular way. The first is governed by the central

nervous system: if someone blows a puff of air into your eyes, for example, you automatically blink. We have no conscious control over these sorts of actions.

The second reason is where people make a deliberate decision to perform a particular act. One of the influencing factors that determines not only their action, but how well they perform it is their motivation: motivation is something that drives behavior. It cannot be used as a singular explanation for why people do things, however, because behavior is also influenced by needs, incentives, expectations, and the presence of conflicting motivations, as well as unconscious factors.

Motivation is usually considered from three aspects: the physiological, the behavioral, and the psycho-social. Here we provide a brief introductory overview, so that you can start to understand *why* an individual user behaves in a particular way. This may be particularly useful if you are designing gaming software, for example, where you want to keep the user engaged by providing the appropriate motivation for them to carry on playing. We will return to the topic of motivation in Chap. 8 to discuss how it plays out in team settings.

4.7.2 Maslow's Hierarchical Theory

Perhaps the best known theory of motivation is Abraham Maslow's (1943) Hierarchy of Needs. The simplest way to visualize the hierarchy is as a pyramid as shown in Fig. 4.19, although Maslow never presented a figure.

At the bottom of the pyramid are the basic, physiological human needs like breathing, food, sleep, and excretion. These are fundamental to human existence. Above that, in order, are safety, love and belonging, and esteem, with self-actualization at the very top. Each level requires that the level below it be mostly satisfied. Maslow also believed in what he called *metamotivation*—the motivation of people for continual personal improvement.

People are orienting towards, and try to satisfy, more than one of these levels at the same time. It is likely that motivations that appear lower in the pyramid will dominate those that appear at a higher level (e.g., eating when you are hungry or going to the bathroom when you feel the need would likely win out over doing something to improve your self-esteem). Whilst there have been many theories of motivation proposed since Maslow's, his basic ideas remain popular.

4.7.3 Extrinsic and Intrinsic Motivation

The basic needs—those towards the bottom of Fig. 4.19—are met for designers and many users, so the higher level needs are the focus of attention from our perspective. In his book, *Drive*, Pink (2009) presents a summary of motivational
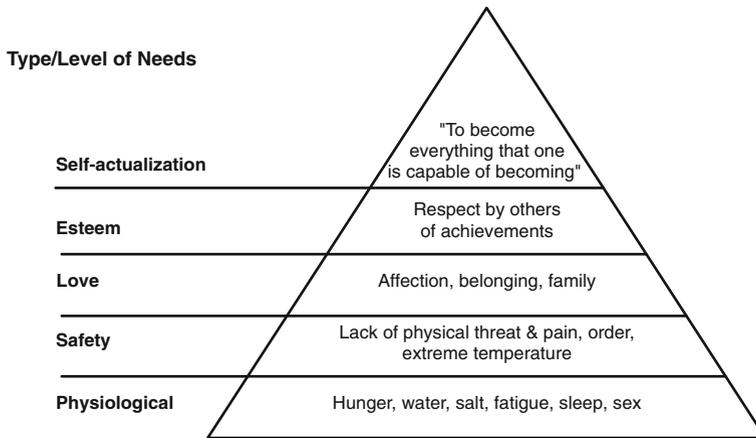
Fig. 4.19 Maslow's hierarchy of needs

psychology research, some of which can be applied to systems design. His work was influenced by that of Deci and others, on Self-Determination Theory (SDT, e.g., Ryan and Deci 2000).

Pink (2009) identifies three versions of motivation:

- Motivation 1.0 is simply about doing what you have to do to survive.
- Motivation 2.0 is associated with the industrial revolution. The underlying assumption is that most people do not want to work, so pressure (both positive and negative) needs to be applied to encourage the right behavior. In other words, extrinsic motivation needs to be supplied, using a carrot and stick approach, such as encouraging work by paying bonuses.
- Motivation 3.0 acknowledges the existence of intrinsic work drives. For many important tasks people want to do the work: they are intrinsically motivated. This includes creative and non-routine work, such as system design, software engineering, and most tasks performed by knowledge workers.

For tasks that are boring or onerous, Pink argues that Motivation 2.0 is still applicable. Where there is little room for creativity, work is generally performed better if the rationale for doing it is provided, the boringness is acknowledged, and people are given some autonomy in how the work is carried out.

For creative and non-routine work, Pink (2009) argues that the keys to high performance are the drives to direct your own life, to extend your abilities, and to live a purposeful life. This sort of work can be hindered by extrinsic motivations. Deci (1971), for example, found that if you paid participants to do a simple puzzle, they were less likely to do it while waiting to take part in the experiment (and hence not being paid), whereas participants who were not paid were more likely to try to solve it while waiting. The argument is that the extrinsic reward focuses

attention on that reward, which facilitates the behavior that the reward encourages, and generally damps other related behavior. Extrinsic rewards may work up to a point, but beyond that, performance can start to decrease. Similarly, if people are focused on the extrinsic reward and it is then withdrawn, this can also lead to a reduction in performance.

The problem lies in achieving the right balance between extrinsic and intrinsic rewards. Pink suggests one way is to increase basic payments and reduce or even eliminate bonus payments. His claim is that people will continue to increase performance if they are intrinsically motivated. In other words, they do the task because they like doing it, or they feel a sense of altruism when they do it (Toms Shoes in California, for example, donates a pair of shoes to needy children for every pair that it sells), or they are given autonomy in what they do (Google, for example, allows employees to spend 20% of their work hours on something new).

Based on his analysis of studies from psychology and behavioral economics, Pink identifies three elements of intrinsic motivation:

- Autonomy—the ability to choose what to work on, how to work on it, who to work with, and when to work on it. If you allow people to make these sorts of decisions they become more productive. Pink argues that autonomy leads to engagement, and engagement leads to mastery. It is not possible for every type of situation, however, so some of the attributes are less mutable than others. If someone is operating a safety critical system in a nuclear power plant, for example, you really *do* want them to follow the rules about how they work.
- Mastery—the desire to understand a process or task and to get better at performing it. Achieving mastery is most enjoyable when the task provides the right level of challenge: too easy and it can become boring; too hard and it can create anxiety. The balancing of knowledge to work requirements is what Vygotsky (e.g., Chaiklin 2003) calls *appropriate scaffolding* or the *Zone of Proximal Development*, and Csíkszentmihályi (1990) calls *flow*. It is not always easy to operationalize these concepts, particularly in social settings, such as a classroom with multiple learners. The desire to achieve mastery can be a powerful drive, although it is not universal. Users who want to become experts will want to understand the task and develop the skills needed to do it well. This drive can also be exploited for boring and difficult tasks as long as the user understands the need for repeated drill exercises, or how the difficult tasks fit into the whole task.
- Purpose—this is really the desire to improve things; most people are at least partly purpose-driven. Under Pink's Motivation 2.0 the purpose is related to the extrinsic motivation. In Motivation 3.0, however, the purpose is related to intrinsic motivation: people know why the task is important, and may even do it free of charge. Emphasizing the goal of the task and its purpose can drastically influence performance. Working for the good of the company, your town, your nation, or the planet in this way can be more motivating than working for money.
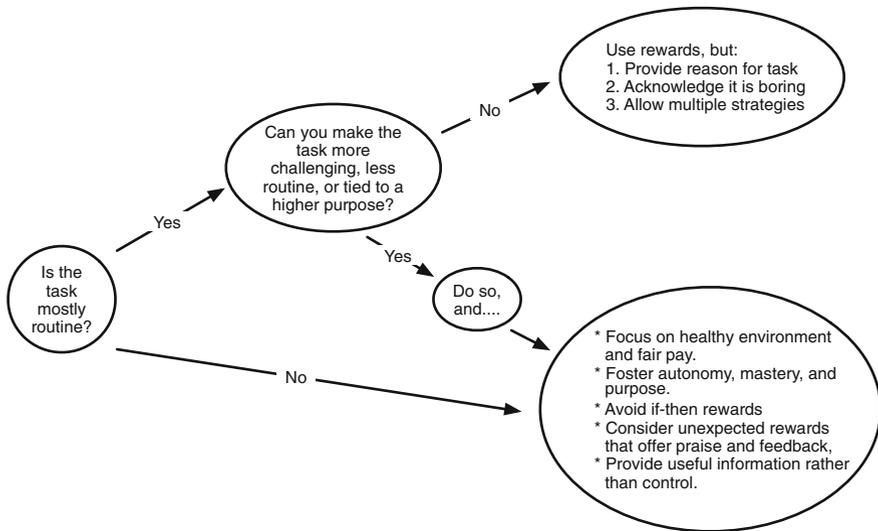
Fig. 4.20 When and how to use rewards based on theories in Pink (2009)

4.7.4 Implications for System Design

How much you need to consider motivation during system design will depend on three things: your particular users, the particular tasks they are doing, and the particular context in which they are working. It is probably best to consider these in reverse order, and think about the context first. If the context is a work setting, then there may be little or nothing that you can do to provide any motivation for your users to do their tasks over and above the motivation they already get from their company. If it is a leisure setting, however, then you will need to look at ways of engaging users in doing their tasks. You could offer virtual monetary rewards, for example, or badges of achievement to show that they had attained a particular level of proficiency at performing the task. Figure 4.20 summarizes this approach.

In terms of the particular tasks they are performing, there are more ways in which you can contribute in work settings. Given what you know about how visual perception and aural perception work, you can now start to think about how to make important items stand out on a crowded screen (using pop-out effects), for example, and how to design audible alarms that can be used to quickly identify a problem (by making them distinctive). In this way you can help your users to improve their task performance, and help them to achieve mastery.

In leisure settings your approach to dealing with the user's tasks may be more perverse, in that you do not want to make it too easy for your users to achieve mastery. You want to make them engage with the game, for example, and spend time and effort to attain any rewards, such as being able to progress to the next level of attainment, on the way to achieving mastery.

Finally, you need to think carefully about your users. People are different: some are driven mostly by extrinsic rewards (particularly money), whereas others may do things for intrinsic rewards, such as the enjoyment or satisfaction they get from doing them.

4.8 Summary

This chapter has provided a more detailed theory of what information users get from their primary perceptual systems. Nearly all users interact through vision and sound, and a deeper understanding of how users get information from these senses provides numerous suggestions for how to improve systems.

Many designers start out with a theory that the eye is like a camera—and so it is in some ways, having a lens and a surface on which the image is registered (the retina). However, as with many metaphors and analogies, the use of the camera metaphor is problematic, and is often called 'naive realism'. It is an inadequate theory of perception.

There are some critical differences between cameras and photography and the eye and visual perception. These differences mean that there is not a simple one-to-one correspondence between the context of a scene and what we perceive. The most important point is that light sensitivity varies across the whole of the retina. It is best in the center of the retina, the fovea, and worst at the edges. The eye thus has to move (saccade) over the visual scene all the time to see the details. Users have to actively do this, and if they do not know where to look or do not know how to interpret an object, they will not see what the designer intends.

If you want to view the eye as a camera, you have to view the eye as a very odd camera. It has a more variable speed film (or sensor) to support wide changes in light than you can buy. It has a poor focusing system (it has to be on the retina to be well developed), and poor quality film over much of the remaining negative (rods). The user has to move the camera to see details. If you view it this way, you probably have a reasonable model of the eye in mind, but a very odd camera.

Hearing is more straightforward. The differences between how the ear works and folk psychology are less drastic. The ear is a little bit more sensitive than most might believe, and can distinguish a wide variety of sounds, but without attention or training some sounds will not be distinguished.

In the future you will see (a) improved displays and (b) more attention given to the integration of types of perception and action. In the first area, of improved displays, you will see continued improvement in the quality of displays. Sound appears to stop improving around 44 kHz sample rates; above that, increasing the quality of the sound is not very noticeable—CDs do not sound a lot different from MP3 files, or so many consumers believe. A similar effect will be found as displays get better, in that the eye can still see more than a display can show. When a display provides the details that the fovea can see across its whole surface, or when

the display adapts to provide more details where the fovea is, the quality will reach a maximum.

In the second area, users integrate their senses—motion, sound, and vision. Current displays are not well integrated with users' movements. This integration will be important for virtual reality (VR). Head-mounted displays move with the user, and most do not compensate for this motion. The users feel their motion with their inner ear, and do not see this motion in the display (or, in some cases, see movement but do not feel it with their inner ear). Over periods of time, sometimes even short periods of times, this non-correspondence between eye and ear can lead to nausea. This mismatch can also appear when an interface is stationary and the user is inside a moving vehicle. Examples include reading books in cars as well as using displays in cars, trucks, and ships. The results are not directly dangerous but they are debilitating and dangerous if the display is necessary for the vehicle. Work is ongoing to understand and ameliorate these effects.

Finally, you will need to think about how to motivate your users to do the tasks they need to do. This will involve balancing their individual needs and aspirations with an appropriate combination of extrinsic and intrinsic rewards if you want to get the best out of them.

4.9 Other Resources

For further descriptions of human perception and performance, it is worth consulting this text:

Boff, K. R. and Lincoln, J. E. (1988). *Engineering data compendium: Human perception and performance*. Wright-Patterson Air Force Base, OH: Harry G. Armstrong Aerospace Medical Research Laboratory. This set of three volumes covers the breadth of what was known about these areas of human behavior at that time. It remains useful because of its breadth.

There are several texts that will tell you more about visual perception. Two we recommend are:

Bruce, V., Green, P. R., and Georgeson, M. A. (2003). *Visual perception: Physiology, psychology and ecology*. Hove, UK: Psychology Press.

Sekuler, R. and Blake, R. (2005). *Perception*. (5th ed). New York, NY: McGraw-Hill.

An excellent introduction to eye movements and their relationship to attention if offered by Goldberg and colleagues:

Goldberg, J. H. and Kotval, X. P. (1999). Computer interface evaluation using eye movements: methods and constructs, *International Journal of Industrial Ergonomics, 24*, 631–645. This provides comparative assessment of measures of eye movement locations and scan paths used for evaluation of interface quality.

You can also get a broad overview of eye gaze and its relationship to attention by looking at Navalpakkam and Churchill's chapter available in a general textbook on HCI methods edited by Olson and Kellogg:

Navalpakkam, V., and Churchill, E. F. (in press). Eyetracking: A brief introduction. In J. S. Olson & W. Kellogg (Eds.), *Ways of knowing, HCI methods*. Heidelberg, Germany: Springer.

To learn more about auditory perception, hearing, look at this text:

Moore, B. C. J. (2013). *An introduction to the psychology of hearing. (6th. Ed).* Bingley, England: Emerald Group Publishing.

Also, a readable text by ex-leader of the band Talking Heads, David Byrne is "How Music Works". It includes a very interesting, and very readable guide to the way that music (as an example of sound) works for the listener. It describes how both analog and digital technology affect the way that we hear sounds, in addition to the context in which the sounds are played.

Byrne, D. (2012). *How Music Works*. Edinburgh, UK: Canongate Books.

To better understand how these basic perceptual capabilities affect how we think, it is worth looking at John Anderson's book on cognitive psychology.

Anderson, J. R. (2009). *Cognitive psychology and its implications* (7th ed.). New York, NY: Worth Publishers. This book includes information about where perception and cognition interact.

The Poynter Institute (http://www.poynter.org) has online tutorials to help designers understand the complexities of color and its use in print and online journalism. These tutorials include page design exercises, which let you experiment with the use of color in magazines, newspapers, and web sites. See also Color, Contrast and Dimension in News Design, https://www.newsu.org/courses/color-news-design (requires free registration).

Clayton Lewis and John Rieman's shareware book: *Task-Centered User Interface Design*, hcibib.org/tcuid, includes discussions about how to visually design an interface to support users. They are familiar with vision, visual cognition, and the application of this knowledge to interface design. Their view helps extend how to apply this knowledge to interface design.

4.10 Exercises

4.1 To study the effects of color on perception you should generate 15–20 stimuli for people to look at on a smartphone. These all have to be items that they have to process in some way, so they could be short paragraphs to read, words

to say out loud or to categorize, or objects to name. The items should be prepared in both black and white and color.

Have a group of users do the task with both the black and white and color objects. Record the time it takes them to do the tasks and how many errors they make. You may have to vary the size of the objects to get these measures to vary.

In addition to reporting what you find, you should indicate what the findings mean about displays and the use of color.

4.2 Choose six icons from popular or not so popular software packages. Also choose six menu items that are words (e.g., Format: Frame, File: Save, Insert: Date, Message: Reply, View: Master Document, Tools: Goal Seek, Edit: Copy) or choose icons from an online training system or training system such as Rosetta Stone.

Ask six to ten people to tell you what they think each icon and each menu item will do or represents. Summarize your results in a table as well as a set of suggestions for how to design interfaces. If you have access to a different population of users, people from different cultures, physical abilities, or ages, run the study on that population as well and compare their responses to your classmates.

4.3 Explain how signal detection theory can be used to analyze web site reading and searching. Based on this analysis, provide three suggestions for your favorite search engine or web site that includes search.

4.4 Redraw the signal detection curve in a big format. Either make the curves big, or put the signal and noise on separate lines to that they are easy to work with. Label the parts of the curves that make up hits, misses, false alarms, and correct rejections.

Redraw the curves with the signal being smaller. How does this affect the setting of the threshold, or how does it influence how to set the threshold for a given ratio of hits to false alarms?

4.5 Choose an online, social web site, such as Facebook, YouTube, or Yelp. Sketch several tasks that can be done by users with the site. Describe the intrinsic and extrinsic motivation(s) for users to perform those tasks. Do the same for an online course site, and for an online game. Note some insights that arise.

References

Bowmaker, J. K., & Dartnall, H. J. A. (1980). Visual pigments of rods and cones in a human retina. *Journal of Physiology, 298*, 501–511.

Catchpole, K., McKeown, J. D., & Withington, D. J. (2004). Localizable auditory warning pulses. *Ergonomics, 47*(7), 748–771.

Chaiklin, S. (2003). The zone of proximal development in Vygotsky's analysis of learning and instruction. In A. Kozulin, B. Gindis, V. Ageyev, & S. Miller (Eds.), *Vygotsky's educational theory and practice in cultural context* (pp. 39–64). Cambridge: Cambridge University.

Csíkszentmihályi, M. (1990). *Flow: The psychology of optimal experience*. New York, NY: Harper and Row.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology, 18*(1), 105–115.

Duchowski, A. T. (2007). *Eye tracking methodology: Theory and practice*. London: Springer.

Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision: The psychology of looking and seeing*. Oxford: Oxford University Press.

Friedrich, M. B. (2008). *Implementierung von schematischen Denkstrategien in einer höheren Programmiersprache: Erweitern und Testen der vorhandenen Resultate durch Erfassen von zusätzlichen Daten und das Erstellen von weiteren Strategien (Implementing diagrammatic reasoning strategies in a high level language: Extending and testing the existing model results by gathering additional data and creating additional strategies)*. Faculty of Information Systems and Applied Computer Science, University of Bamberg, Germany.

Galanter, E. (1962). Contemporary psychophysics. In R. Brown, E. Galanter, E. H. Hess, & G. Mandler (Eds.), *New directions in psychology* (pp. 87–156). New York: Holt, Rinehart & Winston.

Garner, W. R. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. New York, NY: Oxford University Press.

Kennedy, A., & Baccino, T. (1995). The effects of screen refresh rate on editing operations using a computer mouse pointing device. *The Quarterly Journal of Experimental Psychology, 48A*(1), 55–71.

Kubovy, M., & van Valkenburg, D. (2001). Auditory and visual objects. *Cognition, 80*, 97–126.

Maslow, A. H. (1943). A theory of human motivation. *Psychological Review, 50*(4), 370–396.

Moore, R. J., & Churchill, E. F. (2011). Computer interaction analysis: Toward an empirical approach to understanding user practice and eye gaze in GUI-based interaction. *Computer Supported Cooperative Work, 20*(6), 497–528.

Navalpakkam, V., & Churchill, E. F. (in press). Eyetracking: A brief introduction. In J. S. Olson & W. Kellogg (Eds.), *Ways of knowing, HCI methods*. Heidelberg: Springer.

Nielsen, J., & Pernice, K. (2010). *Eyetracking web usability*. Berkeley, CA: New Riders.

Peterson, S. A., & Simon, T. J. (2000). Computational evidence for the subitizing phenomenon as an emergent property of the human cognitive architecture. *Cognitive Science, 24*(1), 93–122.

Pink, D. H. (2009). *Drive*. New York, NY: Riverhead Books.

Reingold, E. M., Loschky, L. C., McConkie, G. W., & Stampe, D. M. (2003). Gaze-contingent multiresolution displays: An integrative review. *Human Factors, 45*(2), 307–328.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*, 68–78.

Sekuler, R., & Blake, R. (2005). *Perception* (5th ed.). New York, NY: McGraw-Hill.

Stevens, S. S., & Newman, E. B. (1936). The localization of actual sources of sound. *American Journal of Psychology, 48*, 297–306.

Strutt, J. W. (Lord Rayleigh, Third Baron of Rayleigh) (1907). On our perception of sound direction. *Philosophical Magazine, 13*, 214–232.

Swets, J. A. (1973). The relative operating characteristic in psychology. *Science, 182*, 990–1000.

Swets, J. A., Tanner, W. P., & Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review, 68*, 301–340.

Treisman, A. M., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12*, 97–136.

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2014). *Engineering psychology and human performance* (4th ed.). Boston, MA: Pearson.

Wickens, T. D. (2002). *Elementary signal detection theory*. Oxford: Oxford University Press.