Chapter 3: Human Factors as HCI Theories

3.1 Human Information Processing

It is easy to understand that HCI design would require an understanding of both computing (software/hardware) and humans. While we will look at the computing aspect in the second part of this book, we take a brief look at some basics in human's side. In chapters 1 and 2, we studied two bodies of knowledge for HCI design, namely high level and abstract principles and more specific guidelines. To practice "user centered" design by following these principles and guidelines, the interface requirements must be often investigated, solicited derived and understood directly from the target users e.g. through focus interviews and surveys. However, it is also possible to obtain a fairly good understanding of the target user from human factors knowledge. Human factors, as the main underlying theory for HCI, can largely be divided into: (1) cognitive science which explains the human's capability and model of conscious processing of high level information and (2) ergonomics which elucidates how raw external stimulation signals are accepted by our five senses, processed up to the pre-attentive level and later acted upon the outer world through the motor organs. Human factors knowledge will particularly help us design HCI in the following ways.

- **Task/Interaction Modeling:** Formulate the steps for how humans might interact to (solve and) carry out a given task/problem and derive the interaction model. Even though a careful HCI designer would not neglect to obtain this model by direct observation of the users themselves, the knowledge in cognitive science will help greatly.

- **Predict, Assess and Evaluate Interactive Behavior:** Understand and predict how humans might react mentally to various ways of information presentation and input solicitation methods as a basis for interface selection. Also, evaluate interaction models and interface implementations and explain/predict their performance and usability

3.1.1 Task Modeling and Human Problem Solving Model

One of the HCI principles we studied was to gain an understanding of the task required for accomplishing the ultimate goal of the interactive system. For instance, a goal of a word processing system might be to produce a nice looking document as easy as possible. In more abstract terms, this whole process of interaction could be viewed as a human attempting to solve a "problem" and applying certain "actions" on "objects" to arrive at a final "solution." Cognitive science has investigated in how humans solve problems, and such a model can help HCI designers analyze the task and base the interaction model or interface structure around it. Thus for a smaller problem of "fixing the font," the action could be a "menu item selection" applied to a "highlighted text." There are several "human problem solving" models that are put forth by a number of researchers, but most of them can be collectively summarized and depicted in Figure 3.1. This problem solving process epitomizes the overall information processing model. First, in the large, the human problem solving or information processing consists of these important parts:

- **Sensation** which senses external information (e.g. visual, aural, haptic) and **Perception** which interprets and extracts basic meanings of the external information. (As a lower level part of the information processing chain (more ergonomic), we take a closer look at these and how they relate to HCI in Section 3.2).

- **Memory** which stores momentary and short information or long term knowledge. This knowledge includes e.g. information about the external world, procedures, rules, relations, schemas, candidates of actions to apply, the current objective (e.g. accomplishing the interactive task successfully) and the plan of action, etc.

- **Decision Maker/Executor** which formulates and revises a "plan," then decides what to do based on the various knowledge in the memory, and acts it out by commanding the motor system (e.g. to click the mouse left button).

Figure 3.1(b) shows the overall process in a flow chart. Once a problem is defined and decided that it is to be solved (simply by the user's intention), it is established as the top "goal." Then a hierarchical plan (Figure 3.2) is formulated by refining the goal into a number of sub-goals. A number of actions or sub-tasks are identified in hopes of solving the individual sub-goals considering the external situation. By enacting the series of these "sub-tasks" to solve the sub-goals, the top goal is eventually accomplished. Note that enacting the sub-tasks does not guarantee their successful completion (i.e. they may fail). Thus the whole process is repeated by observing the resulting situation and revising and restoring the plan.

[Figure 3.1]: (a) The overall human problem solving model and proccess and (b) a more detailed view of the "Decision Maker/Executor.

Figure 3.2 shows an example of a hierarchical task plan (equivalent to hierarchical goal structure) illustrating how the simple task of "changing the font of a text" could be solved, and in another words, what kind of basic tasks would be needed. Note that in a general hierarchical task model, certain sub-tasks need to be applied in series and some may be concurrently. One can readily appreciate from the simple example in Figure 3.2 how an interactive task model can be hierarchically refined and can serve as a basis of the interface structure. Note, based on this model, we could "select" interfaces to realize each sub-task in the bottom of the hierarchy, which illustrates the crux of the HCI design process. The interaction model must represent as much as possible what the user has in mind and expects what one would do (the mental model) in order to accomplish the overall task. This way, the user will be "in tune" with the resulting interactive application. The interface selection should be done based on ergonomics, user preference, other requirements or constraints for high benefit (usability and performance) to cost. Finally, the subtask structure can lend itself to the menu structure, and the actions and objects to which the actions apply can serve as the basis for an object-class diagram (for an object oriented interactive software implementation).

[Figure 3.2] An example of a hierachrical task model of changning a font for a short text. Note that a specific interface may be chosen to accomplish the sub-tasks in the bottom.

3.1.2 Human Reaction and Prediction of Cognitive Performance

We can also to some degree predict how humans will react and perform to a particular human interface design. We can consider two aspects to human performance, one that is cognitive and the other ergonomic. In this section, we focus on the former.

Norman spoke of the "Gulf of Execution/Evaluation" in which how users will can be left bewildered (and will not perform very well) when an interactive system does not offer certain actions or does not result in a state as expected by the user [22]. Such a phenomenon would be a result of an interface based on an ill-modeled interaction. A user, when solving a problem or using an interactive system to do so, will first form a mental model which is mostly equivalent to the hierarchical "action" plan for the task (see Section 3.1.1). The mismatch between the user's mental model and the task model employed by the interactive system creates the "Gulf." On the other hand, when the task model and interface structure of the interactive system maps well to the expected mental model of the user, the task performance will be very fluid.

[Figure 3.3] Gulf of executiona and evaluation: the gap between the expected and actual.

Memory capacity also influences the interactive performance greatly. As shown in Figure 3.1, there are largely two types of memories in the human cognitive system, the short term and the long term. The short term memory is also sometimes known as the working memory in the sense that it contains (changing) memory elements meaningful for the task at hand (or chunks). Humans are known to remember about 8 chunks of memory lasting only a very short amount of time [20]. This means an interface cannot rely on the human's short term memory beyond this capacity for fast operation. Imagine an interface with a large number of options or menu items. The user would have to rescan the available options number of times to make the final selection. In an on-line purchasing system, the user might not be able to remember all the relevant information such as items purchased, delivery options, credit card chosen, billing address, usage of discount cards and etc. thus such information will have to be presented to the user time to time for refreshing one's memory and making sure no errors are made.

[Figure 3.4] A snap shot of an online shopping process that does not display superfluous user status that resultantly brings about anxiety, uncertainty and erroneous response.

Retrieving information from the long term memory is a difficult and relatively time consuming task. Therefore, if an interactive system (e.g. targeted even for experts) requires expert-level knowledge, it needs to be displayed as to at least elicit "recognition" (among a number of options) of it rather than completely rely on “recall” from scratch.

Memory related performance issues are also important in multitasking. Many modern computing environments offer multitasking environments. It is known that when the user switches the task from one to the other, a “context switch” occurs in the brain, which means the working memory content is replaced (and stored back into the long term memory) with chunks relevant (such as the state of the task up to that moment) for the switched task. This process can bring about overall task performance degradation in many respects [19]. For an individual application to help itself in its use during multitasking, it can assist the user’s context switch process by capturing the context information during its suspension, and by later displaying, reminding and highlighting the information upon resumption (Figure 3.5).

[Figure 3.5] Reminding the user of the context for multitasking for fast application switching (top part of the figure).

3.1.2.1 Predictive Performance Assessment: GOMS [8]

Many important cognitive activities have been analyzed in terms of their typical approximate process time, e.g. for single chunk retrieval from the short term memory, encoding (memorizing) of an information into the long term memory, responding to a visual stimuli and interpreting its content, and etc. [2][25][28]. Based on these figures and a task sequence model, one might be able to quantitatively estimate the time taken to complete a given task, and therefore make an evaluation with regards to the original performance requirements. Tables 3.1 and 3.2 illustrate such an example based on the framework called GOMS [8].

The GOMS evaluation methodology starts by the same hierarchical task modeling we have explained in Section 3.1. Once a sequence of sub-tasks is derived, one might map a specific operator in Table 3.1 (or in other words interface) to each of the sub-task. With the pre-established performance measures (Table 3.1), the total time of task performance can be easily calculated by summing the task times of the whole set of sub-tasks. Different operator mappings can be tried comparatively in terms of their performance. The original GOMS model was developed mainly for the desktop computing environment with performance figures for mouse clicks, keyboard input, hand movement, and mental operators (Table 3.1). Even though this model was created nearly 30 years ago, the figures are still amazingly valid (while computer technologies have advanced much since then, humans capabilities have remained mostly the same). GOMS for other computing environments have been proposed as well [30].

Table 3.1 shows different performance measures for various task operators or interfaces [8]. Table 3.2 shows two "designs" of the main task of "file deletion." Each design is decomposed in a slightly different manner and operators mapped to the individual subtasks, resulting in different total times of operation (the first in 4.8 seconds and the second in 2.6 seconds). GOMS is quite simple in that it can only evaluate in terms of the task performance, while there are many other criteria by which an HCI design should be evaluated. Moreover, among the operators, the "Mental" operator, approximates the time taken for "momentary thought or memory retrieval" in between motor tasks (like mouse clicks). Obviously, there can be some inaccuracies introduced in the use of the Mental operators during the interaction modeling process[1]

[Table 3.1] Estimates of time taken for typical desktop computer operations from GOMS [8].

Type of Operation	Time Estimate
K: Keyboard input	Expert: 0.12 sec
	Average: 0.20 sec
	Novice: 1.2 sec
T(n): Type n characters	280*n msec
P: Point with mouse to something on the display	1100 msec
B: Press or release mouse button	100 msec
BB: Click a mouse button (Press and release)	200 msec
H: Home hands, either to the keyboard or mouse	400 msec

[1] The developers of GOMS do outline a strategy for when to properly use the Mental operators for a correct task modeling [8].

M: Thinking what to do (Mental operator)	1200 msec (can change)
W(t): Waiting for the system (to respond)	t msec

[Table 3.2]: Estimates of time taken for two task models of "Delete a File." Design 2 is the "expert" version that uses a "hot key" [8]. The total time is computed by simply adding the corresponding figures in Table 3.1.

<table>
<tr><td colspan="4">"Deleting a File"</td></tr>
<tr><td colspan="2">Design 1</td><td colspan="2">Design 2</td></tr>
<tr><td>1. Point to file icon</td><td>P</td><td>1. Point to file icon</td><td>P</td></tr>
<tr><td>2. Click mouse button</td><td>BB</td><td>2. Click mouse button</td><td>BB</td></tr>
<tr><td>3. Point to file menu</td><td>P</td><td>3. Move hand to keyboard</td><td>M</td></tr>
<tr><td>4. Press and hold mouse button</td><td>B</td><td>4. Hit command key command-T</td><td>KK</td></tr>
<tr><td>5. Point to DELETE item</td><td>P</td><td>5. Move hand back to mouse</td><td>H</td></tr>
<tr><td>6. Release mouse button</td><td>B</td><td></td><td></td></tr>
<tr><td>7. Point to original window</td><td>P</td><td></td><td></td></tr>
<tr><td colspan="2">Total Time = 4.8 sec.</td><td colspan="2">Total Time = 2.66 sec.</td></tr>
</table>

3.2 Sensation and Perception of Information

The previous section explained the value and usage of the knowledge of cognitive and high level information processing to HCI design and we now shift our focus to raw information processing. First we look at the input (to human sensory system) side. Humans are known to have at least five senses. Among them, those that would be relevant to HCI (at least for now) are the modalities of visual, aural, haptic (force feedback) and tactile. Taking external stimulation or raw sensory information (sometimes computer generated) and processing it for perception is the first part in any human computer interaction. Naturally, the information must be supplied in a fashion that is amenable to human's consumption, that is, within the bounds of human's perceptual capabilities.

Another aspect of sensation and perception is **"attention,"** that is, how to make the user selectively (consciously or otherwise) tune to a particular part of the information or stimulation. Highly attentive information can be used for alerts, reminders, highlighting of prioritized/structured information, guidance, etc. Note that attention must occur and be modulated within **"awareness"** of the larger task(s). While we might tune to certain important information, we often still need to have an understanding, albeit approximate, of the other activities or concurrent tasks, such as in multitasking or parallel processing of information.

In what follows, we examine the process of sensation and perception in the four major modalities and human's capabilities in this regard. Just as cognitive science was useful to interaction and task modeling, this knowledge is essential in sound interface selection and design.

3.2.1 Visual

Visual modality is by far the most important information medium. Over 40% of the human brain is said to be involved with the visual information processing. As already mentioned, the visual interface design and display system parameters will have to conform to the capacity of the human visual system and characteristics. In this section, we review some of the important properties of the human visual system and the implication to interface design. First we take a look at a typical visual interaction situation as shown in Figure 3.6.

[Figure 3-6] A user viewing a display system. The shaded area illustrates the horizontal field of view (shown to be much less than the actual for illustration purpose) while the dashed line the same as offered by the display. The display offers different field of view depending on the viewing distance (dotted line in the middle). The oval shape in the display represents the approximate area for which high details are perceived through the corresponding the foveal area in the user eyes.

3.2.1.1 Visual and display parameters

- Human's Field of View (FOV) – This is the angle subtended by the visible area by the human user in horizontal or vertical direction. The shaded area in Figure 3.6 illustrates the horizontal field of view. Human's FOV is nearly 180 degrees in both horizontal and vertical directions.

- Viewing Distance – The perpendicular distance to the surface of the display. Viewing distance (dotted line in Figure 3.6) may change with user movements. However, one might be able to define a nominal and typical viewing distance for a given task or operating environment.

- Display Field of View – This is the angle subtended by the display area from a particular viewing distance. Note that for a same fixed display area, the display FOV will be different at different viewing distances. In Figure 3.6, the display FOV is denoted with the dashed line. The display offers different fields of view depending on the viewing distance (dotted line in the middle).

- Pixel – A display system is typically composed as an array of small rectangular areas called the pixel.

- Display Resolution – The number of pixels in horizontal and vertical direction for a fixed area.

- Visual Acuity – In effect, the resolution perceivable by the human eye from a fixed distance. This is also synonymous to the power of sight which is different for different people and age groups.

[Figure 3.7] The display system parameters: display size, resolution, pixel determined by the user's visual acuity and viewing location.

These human visual and display parameters need to be matched as much as possible to provide a comfortable and effective visual display environment, for instance, display FOV to human FOV, display resolution/ object size to visual acuity and so forth. Note that the display FOV is more important than the absolute size of the display. A distant large display can have the same display FOV as a close small display, even though it may incur different viewing experiences. If possible, it is desirable to choose the most economical display, not necessarily the biggest or the one with highest resolution, with respect to the requirement of the task and the typical user characteristics.

3.2.1.2 Detail and peripheral vision

The human eye contains two types of cells that react to light intensities in different ways. The "Cones" are responsible for color and detail recognition are distributed heavily in the center of the retina (back of the eyeball), which subtends about 5 degrees in the human's FOV and roughly establishes the area of focus. The oval region in Figure 3.6 shows the corresponding region in the display for which details can be perceived through these cells. On the other hand, the "Rods" are distributed mainly in the periphery of the retina and are responsible for motion detection and less detailed peripheral vision. While details may not be sensed, it contributes to our awareness of the surrounding environment.

Differently from that of the human, most displays have uniform resolution. However, if the object details can be adjusted depending on where the user is looking at or based on what the user may be interested in (Figure 3.8(a)), the overall rendering of the image can be made more economical or to doubly emphasize certain object relative to others in its neighborhood (e.g. detail contrast). We may assess the utility of a large, very high resolution display systems such the one shown in Figure 3.9. From a nominal viewing distance only a small portion of the large display will correspond to the foveal area. Thus, while viewing one portion, the other major parts of the large display resolution could go to "waste" (unless used by multiple users at once as in an IMAX theater). Thus, it can be argued that it is more economical to just use a smaller high resolution display used from a close distance. Interestingly, Microsoft Research recently introduced a display system called the "Illumiroom" in which a high resolution display is used in the middle and a wide low resolution projection and peripheral display provides high immersion (Figure 3.8(b)).

[Figure 3.8] (a) An ideal display would provide relatively higher resolution in the area of user's focus, (b) A large immersive display realized by a high resolution monitor in the middle and lower resolution projection in the periphery (Microsoft Research Illumiroom [13]).

Figure [3.9] A large tiled high resolution display[2] [21]. Is it really worth the cost?

3.2.1.3 Color, Brightness and Contrast

Another important properties and attributes of visual quality are the brightness, color and their contrast.

- Brightness – The amount of light energy emitted by the object (or as perceived by the human).

- Color – Human response to different wavelengths of light, namely for those corresponding to red, green, blue and their mixtures. A color can be specified by the composure by the amounts contributed by the three fundamental colors or also by hue (particular wavelength), saturation (relative difference in the major wavelength and the rest in the light) and brightness/value (total amount of the light energy).

[Figure 3.10] Color specification by hue (particular/dominant wavelength), saturation (relative difference in the major wavelength and the rest) and value/brightness (total amount of the light energy).

---

[2] Visbox, http://www.visbox.com/imgs/granite2.html

- Contrast– Relative difference in brightness or color between two visual objects. Contrast in brightness is measured in terms of the difference or ratio of the amounts of light energies between (two or more) objects. The recommended ratio of the foreground to background brightness contrast is at least 3:1. Color contrast is defined in terms of differences or ratio in the dimensions of hue and saturation. It is said that the brightness contrast is more effective for detail perception than the color contrast (Figure 3-11).

[Figure 3-11] Coding of information in a map (e.g. temperature levels) using contrast in brightness (left) and color (right) [10].

3.2.1.4 Pre-attentive Features and High-level Diagrammatic Semantics

Detail, color, brightness, and contrast are all very low level raw visual properties. Before all these low level part features are finally consolidated for conscious recognition (of a larger object) through the visual information processing pipeline, "pre-attentive" features might be used to attract our attention. Pre-attentive features are composite, primitive and intermediate visual elements that are automatically recognized before entering our consciousness, typically within 10 ms after entering the sensory system [31]. These features may rely on the relative difference in color, size, shape, orientation, depth, textures, motion, and others. Figure 3.12 shows several examples and how they can be used collectively to form and design effective "icons."

At a more conscious level, humans may universally recognize certain high level complex geometric shapes and properties as a whole and understand the underlying concepts. Figure 3.13 shows examples of such universally accepted (across different cultures) geometric diagrams with the connotation of e.g. connection/relation, dependency, causality, inclusion, hierarchy/structure, flow/process, etc.

[Figure 3.12] Examples of pre-attentive features for attention focus based on differences in size, shape, and orientation (left) and application to icon design (right) [21, 28].

[Figure 3.13] Examples of diagrams/shapes/objects/figures with universal semantics [31].

3.2.2 Aural

Next to the visual, the aural modality (sound) is perhaps the most prevalent mode for information feedback. The actual form of sound feedback can be grossly divided into three types: (1) simple beep-like sounds, (2) short symbolic sound bytes known as the earcons (e.g. the paper crunching sound when a file is inserted into the trashcan for deletion), (3) relatively longer "as is" sound feedback that is replayed from recordings or synthesis. As we did for the visual modality, we will first go over some important parameters of the human aural and the corresponding aural display parameters.

3.2.2.1 Aural and Aural Display Parameters

- **Intensity** (amplitude) refers to the amount of sound energy and is synonymous to the more familiar term, “volume.” Intensity is often measured in the units of decibels (dB), a logarithmic scale of sound energy. 0 dB corresponds to the lowest level of audible sound and about 130 dB, the highest. It is instructive to know the dB levels of different sounds as a guideline in setting the nominal volume for the sound feedback (Table 3.3).

[Table 3.3] Examples of different sounds and their typical intensity levels in decibels.

Intensity in dB	Description
0	Weakest sound audible
30	Whisper
50	Office environment
60	Normal conversation
110	Rock band
130	Pain threshold

- Sound can be viewed as containing or being composed of a number of sinusoidal waves with different **frequencies** and corresponding amplitudes. The dominant frequency components determine various characteristics of sounds such as the pitch (e.g. low or high key), timbre (e.g. which instrument) and even directionality (where is the sound coming from?). Humans can hear sound waves with frequency values between about 20 and 20,000 Hz [24].

- **Phase** refers to the time differences among sound waves that emanate from the same source. Phase differences occur, for example, because our left and right ears may have slightly different distances to the sound source and as such phase differences are known to also contribute to the perception of spatialized sound such as stereo.

When using aural feedback, it is important for the designer to set these fundamental parameters properly. A general recommendation is that the sound signal should be between 50 and 5000 Hz and composed of at least four prominent harmonic frequency components (frequencies that are integer multiples of one another), each within the range of 1000 to 4000 Hz [6]. Aural feedback is more commonly used in intermittent alarms. However, overly loud (i.e. needlessly high amplitude) alarms are known to rather startle the user and lower the usability. Instead, other techniques can be used to draw attention to and convey urgency by the aural feedback such as repetition, variations in frequency and volume, gradual and aural contrast to the background ambient sound (e.g. in amplitude and frequency).

3.2.2.2 Other Characteristics of Sound as Interaction Feedback

We further point out few differences of aural feedback from the visual. First, sound is effectively omni-directional. For this reason, sound is the most often used to attract and direct user's attention, but as already mentioned, it can also be a nuisance as a task interrupter (e.g. a momentary loss of context) by the startle effect. Making use of contrast is possible with sound as well. For instance, auditory feedback would require 15~ 30 dB difference to the ambient noise to be hear effectively. Differentiated frequency components can be used to convey certain information.

Continuous sound is somewhat more subject to becoming habituated (e.g. elevator background music) than stimulation with other modalities. In general, only one aural content can be interpreted at a time. That is, it is difficult to make out the aural content when the sound is jumbled/masked with multiple sources. Humans do possess an ability to tune to a particular part of the sound (e.g. string section in a symphony), however this requires much concentration and effort.

3.2.2.3 Aural modality as Input Method

So far, the aural modality was explained only in the context of passive feedback. As for using it actively as means for input to interactive systems, two major methods are: (1) key word recognition and (2) natural language understanding.

Isolated word recognition technology (for enacting simple commands) has become very robust lately. In most cases, it still requires speaker-specific training or a relatively quiet background. Another related difficulty with voice input is the “segmentation” problem, i.e. how to segment out, from a stream of continuous voice input or background noise, the portion that corresponds to the actual command. As such, many voice input systems operate in an explicit mode or state. For example the user has to press a button to activate the voice recognition (and enter into the recognition mode/state), and then speak the command into the microphone (this also relieves the computational burden of having to run the voice recognition process in the background if the system did not know when the command was to be heard). This "switching to the voice command" mode is still quite a nuisance to the ordinary user. Thus, voice input is much more effective in situations e.g. where hands are totally occupied, or modes are not necessary because there is very little background noise or because there is no mixture of conversation with the voice commands.

Machine understanding of long sentences and natural language based commands are still very computationally difficult and demanding. While not quite practical for everyday user interface input methods, language understanding technology is advancing fast as demonstrated recently by Apple's SIRI [34] and IBM Watson [36] where high quality natural language understanding services are offered by the cloud. Captured segments of voice/text input sentences can be sent to these cloud servers for very fast and near real time response. With the spread of smart client media devices which might be computationally light yet equipped with a sleuth of sensors, such a cloud based natural language interaction (combined with intelligence) will revolutionize the way we interact in a near future.

[Figure 3.14] A high quality natural language based interaction through the cloud. The smart media client devices would send the captured sentence (in voice or text) and a correct and intelligent response is given back in real time.

3.2.3 Tactile and Haptic

Interfaces with tactile and haptic feedback, while not very wide spread yet, are starting to appear in limited forms. To be precise, the term "haptic" is defined to be the modality that takes advantage of touch by applying forces, vibrations, or motions to the user [38]. Thus the haptic refers both the sensation of force feedback and touch (tactile). For convenience, we will use the term "haptic" to refer to the modality for sensing force and kinesthetic feedback through our joints and muscles (even though any force feedback practically requires contact through the skin) and "tactile" for sensing different types of touch (e.g. texture, light pressure / contact, pain, vibration and even temperature) through our skin.

3.2.3.1 Tactile and Tactile Display Parameters

- **Tactile resolution** – The skin sensitivity to physical objects are different over the human body. The fingertip is one of the most sensitive areas and frequently used for HCI purpose. It has been known the fingertip can sense up to 40 μm sized objects [3, 27].

- **Vibration frequency** – Rapid movement such as vibration is mostly sensed by the Pacinian corpuscle which is known to have the signal response range of 100 to 300 Hz. Vibration frequency of about 250 Hz is said to be the optimal for comfortable perception [36].

- **Pressure threshold** – The lightest amount of pressure humans can sense is said to be about 1000 N/m^2. For a fingertip, this amounts to about 0.02 N for the fingertip area [37]. The maximum threshold is difficult to measure, when the force/torque gets large enough, the kinesthetic senses start to operate and this threshold will depend much on the physical condition of the user (e.g. strong vs. weak user).

As mentioned above, there are many types of tactile stimulation, such as texture, pressure, vibration, and even temperature. For the purpose of HCI, the following parameters are deemed important and same goes for the display system providing the tactile based feedback. Physical tactile sensation is felt by a combination of skin cells and nerves tuned for particular types of stimulations, e.g. the Meissner's corpuscle for slight pressure or slow pushing (stimulation signal frequency of (3~40Hz), Merckle cells for flutter and textured/protrusion surface (0.3~3Hz), Pacinian corpuscle for more rapid vibratory stimulation (10~500Hz), and Ruffini endings for skin stretch (Figure 3.15

[Figure 3.15] Cells and nerves in the skin[3].

[3] Proprioception, Intl. Encyclopedia of Rehabilitation, http://cirrie.buffalo.edu/encyclopedia/en/article/337/

While there are many research prototype or commercial tactile display devices, the most practical one is the vibration motors, mostly applied in a single actuator configuration. Most vibration motors do not offer separate controllability for its amplitude and frequency. In addition, most vibrators are not in direct contact with the stimulation target (e.g. the hand) making the signal somewhat muffled through the casing. Thus additional care is needed to set the right parameter values for the best effects under the circumstance.

Another way to realize vibratory tactile display is to use thin and light piezo-electric materials which exhibit vibration responses according to the amounts of electric potential supplied. Due to its flat form factor, they may be, for instance, embedded into flat touch screens. Sometimes sound speakers can be used to generate indirect vibratory feedback with high controllability (responding to wide ranges of amplitude and frequency signals).

[Figure 3.16] Left: Various actuators used for tactile feedback (a) miniature speaker, (b) miniature electro-magnet/latch, (c) piezo-electric strip, (d) micro-vibratory motors, Right: (e) tactile array with multiple actuators [5].

3.2.3.2 Haptic and Haptic Display Parameters

Along with the tactile feedback, haptic feedback adds a more apparent physical dimension to interaction. Force feedback and movement is felt by the cells and nerves in our muscle and joints. For instance, the muscle spindle/tendon takes the inertial load and Pacinian/Ruffini/Golgi senses the joint movements, pressure and torque [36]. The activation force for the joints is between 0.5 to 2.5 mN [14]. However, we can understand that this range would vary according to the user's age, gender, strength, size, weight, and so forth. Note that haptic devices are both input and output device at the same time (we briefly discuss this issue of haptic input in the next section in the context of human body ergonomics).

The simplest form of a haptic device is a simple electro-magnetic latch used a lot for game controllers. It generates a sudden inertial movement and slowly repositions itself for repeated usages. Normally, the user holds on to the device and inertial forces are delivered in the direction relative to the game controller. Such a device is not appropriate for fast occurring interaction (e.g. successive gun shots) or for display a continuously sustained force (e.g. leaning against a wall).

More complicated haptic devices are in the form of a robotic kinematic chain, either fixed on the ground or worn on the body. As a kinematic chain, they offer higher degrees of freedom and finer force control (Figure 3.17). For the grounded device, the user interacts with the tip of the robotic chain through which a force feedback is delivered. The sensors in the joints of the device make it possible to track the tip (interaction point) within the 3D operating space. Body worn device transfers force with its mechanism directly attached to the body with the similar control structure.

[Figure 3.17] Two types of haptic systems: (a) grounded[4] and (b) body-worn[5].

Important haptic display parameters are the **degrees of freedom** (in how many directions can force or torque be displayed), **force range** (e.g. should be at least greater than 0.5mN), (3) **operating/interaction range** (how much movement is allowed through the device), and (4) **stability** (how stable the supplied force is felt). Stability is in fact a by-product of the proper "sampling period," which refers to the time taken to sense the current amount of force at the interaction point, determine whether the target value has been reached and reinforce it (a process that repeats until a target equilibrium force is reached at the interaction point). The ideal sampling period is about 1,000 Hz and when the sampling period falls under certain value, the robotic mechanism exhibits instability (e.g. exhibited in the form of vibration), and thus lower usability. The dilemma is that providing high sampling rate requires heavy computation load, not only in updating the output force but also in physical simulation (e.g. to check if the 3D cursor has hit any virtual object). Therefore, a careful "satisficing" solution is needed to balance the level of the haptic device performance and user experience.

---

4 Phantom, http://www.sensable.com/

5 ExoHand, http://www.festo.com/

[Figure 3.18] Important parameters for a haptic display system.

In general, due to their mechanical nature, "robotic" haptic devices are not so practical yet. They tend to be heavy, clunky, dangerous, and take up a large volume. The cost is very high with often only small operating range, force range or limited degrees of freedom. In many cases, simpler devices, such as one directional latches or vibrators, are used in combination with the visual and aural feedback to enrich the user experience.

3.2.4. Multimodal Interaction

Conventional interfaces have been mostly visually oriented. However, for various reasons "multimodal" interfaces are gaining popularity with the ubiquity of multimedia devices. By employing more than one modality, interfaces can become more effective in number of ways depending on how they are configured [27]. Here are some representative examples.

- **Complementary** – Different modalities can assume different roles and complementary to achieving certain interaction objective. For example, an aural feedback can signify the arrival of a phone call, while the visual displays the caller's name.

- **Redundant** – Different modality input methods or feedback can be used to ensure a reliable achievement of the interaction objective. For instance, a phone call "ring" can be a simultaneously aural and tactile to strengthen the pick-up probability.

- **Alternative** – By providing alternative ways to interact, users have more choices in the way they can interact. For instance, a phone call can be made either by touching a button or by speaking the callee's name, and thereby promote convenience and usability.

For multimodal interfaces to be effective, each feedback must be properly synchronized and consistent in its representation. For instance, to signify a button touch, the visual highlighting and beep sound effect must occur within a short time (e.g. less than 200ms) to be recognized as one consistent event. The representation must be coordinated between the two; in the above example, if there is one highlighting, then there should also be one corresponding beep. When inconsistent, the interpretation of the feedback can be confusing or only the dominant modality will be recognized.

3.3 Hunan Body Ergonomics (Motor Capabilities)

So far, we have mostly talked about human's cognitive and perceptual capabilities and how display or input systems must be configured to match them. In this section, we briefly look at ergonomics aspect. To be precise, ergonomics is a discipline focused on making products and interfaces comfortable and efficient. Thus, broadly speaking, it encompasses mental and perceptual issues, although in this book, we restrict the term to mean ways to design interfaces or interaction devices for comfort and high performance according to the physical mechanics of the human body. For HCI, we focus on humans motor capabilities which are used to make input interaction. We start with the Fitt's Law and human motor control.

3.3.1 Fitt's Law [9]

Fitt's Law is a model of human movement which predicts the time required to rapidly move to a target area, as a function of the distance to the target and the size of the target. The Movement task's difficulty index can be quantified in terms of the required information amount, i.e. in the number of bits.

[Figure 3.19] Illustration of Fitt's Law [17].

From the main equation in Figure 3.19, the actual time to complete the movement task is predicted using a simple linear equation, where movement time, *MT,* is a linear function of *ID*.

$$MT = A * ID + B,$$

where A and B are coefficients specific to a given task.

Thus, to reiterate, *ID* represents an abstract notion of difficulty of the task, while *MT* is an actual prediction value for a particular task. The values for coefficients *A* and *B* are obtained by taking samples of the performance and mathematically deriving them by regression (Figure 3.20).

[Figure 3.20] Deriving the actual movement time by fitting based on samples of performance data [17].

Note that the original Fitt's Law was created for interaction with everyday objects (in the context of operation in factory assembly lines) rather than for computer interfaces. Researchers have applied the concept of Fitt's Law to computer interfaces and have found the same principle applies. For instance, as shown in Figure 3.21, the task of "dragganig of an icon into a trashcan icon" using a mouse can be assessed using the Fitt's Law [18]. Many other computer interactive tasks can be modeled similarly and several revised Fitt's Laws (e.g. for desktop computer interface, mobile interface) have been derived as well [11]

[Figure 3.21] Applying Fitt's Law to a computer interface (dragging a file icon into the trashcan icon) [18].

3.3.2 Motor control

Perhaps the most prevalent of form of input is made by the movements of our arms, hands and fingers for keyboard and mouse input. Berard et al. have reported that there was a significant drop in human motor control performance below certain spatial resolution threshold [3]. For instance, while the actual performance is dependent on the form factor of the device used and the mode of operation, the mouse is operable with up to a spatial resolution in the order of thousands of dpi (dots per inch) or ~0.020 mm, and the 3D stylus in the hundreds.

In addition to discrete event input methods (e.g. buttons), modern user interfaces make heavy use of continuous input methods in the 2D space (e.g. mouse, touchscreen) and increasingly in the 3D (e.g. haptic, Wii-mote). While the human capabilities will determine the achievable accuracy in such input methods, the control-display ratio (C/D ratio) is often adjusted. C/D ratio refers to the ratio of the movement in the control device (e.g. mouse) to that in the display (e.g. cursor). If the C/D ratio is low, the sensitivity of the control is high and, therefore, travel time across the di splay will be fast. If the C/D ratio is high, sensitivity is low and, therefore, fine adjust time will be relatively fast.

Obviously, humans will exhibit different motor control performances with different devices as already demonstrated with the two types of device mentioned above (e.g. mouse vs. 3D stylus). The mouse and 3D stylus, for instance, belongs to what is called the "isometric" devices where the movement of the device directly translates to the movement in the display (or virtual space). Non-isometric devices are those that control the movement in the display in principle with something else such as force, thus possibly with no movement input at all.

Control accuracy for touch interface presents a different problem. Despite our fine motor control capability of sub-mm performance and recent touchscreens offer higher than 4096 dpi resolution, it is the size of the fingertip contact (unless using a stylus pen), 0.3~0.7 cm, that makes it hard to make selection for relatively small objects. Even larger objects, once selected, are not easy to control if the touch screen is held by another hand or arm (e.g. unstable). .

3.4 Others

There are many cognitive, perceptual and ergonomic issues that have been left out. Due to the limited scope of this book, we only identify some of them for the readers to investigate further.

- Learning and adaptation
- Modalities other than the "big three (visual/aural/haptic-tactile)" such as gestures, facial expression, brain waves, physiological signals (such as electromyogram, heart rate, skin conductance), gaze, etc.
- Aesthetics and emotion
- Multitasking

3.5 Summary

In this section, we reviewed the essence of human factors, including sensation, perception, information processing and the Fitt's Law, as the foremost underlying theory for human computer interaction design. By the very principle of "Know Thy User," it is adamant that the HCI designer has the basic understanding of these areas so that any interface will suit the user's most basic mental, perceptual and ergonomic capabilities. We can also readily see that many HCI principles discussed in early part of the book naturally derives from these underlying theories.

Reference

[1] Adams, Richard J., Moreyra, Manuel R., and Hannaford, Blake. "Stability and performance of haptic displays: Theory and experiments." *Proceedings ASME International Mechanical Engineering Congress and Exhibition*. (1998).

[2] Anderson, J. R., et al. (2004) An integrated theory of the mind. Psychological Review, 111(4). 1036-1060

[3] Bérard, François, Wang, Guangyu, and Cooperstock, Jeremy R. "On the limits of the human motor control precision: the search for a device's human resolution." *Human-Computer Interaction–INTERACT*. Springer Berlin Heidelberg, (2011). 107-122.

[4] Bolanowski Jr, Stanley J., Gescheider, G. A., Verrillo, R. T., and Checkosky, C. M. "Four channels mediate the mechanical aspects of touch." *The Journal of the Acoustical society of America* 84 (1988): 1680.

[5] KU Leuven (PMA Division) "Tactile Feedback." https://www.mech.kuleuven.be/en/pma/research/ras/researchtopics/tactfb.html, (2014).

[6] Bregman, Albert S. *Auditory scene analysis: The perceptual organization of sound*. A Bradford Book, (1994).

[7] Card, Stuart K., Moran, Thomas P., and Newell, Allen. *The psychology of human computer interaction.* Lawrence Erlbaum, (1983).

[8] Card, Stuart K., Moran, Thomas P., and Newell, Allen. "The model human processor: An engineering model of human performance." *Handbook of Human Perception 2*, (1986).

[9] Fitts, Paul M. "The information capacity of the human motor system in controlling the amplitude of movement." *Journal of experimental psychology* 47.6 (1954): 381.

[10] Hemer, Mark A., Fan, Yalin, Mori, Nobuhito, Semedo Alvaro, and Wang, Xiaolan L. "Projected changes in wave climate from a multi-model ensemble mark", *Nature Climate Change 3*, (2013): 471-476.

[11] I. Scott MacKenzie (1992). Fitts' law as a research and design tool in human computer interaction. Human Computer Interaction, volume 7, 1992, pp. 91?139.

[12] ISO, "ISO 9241-3, Ergonomic requirements for office work with visual display terminals (VDTs) Part 3: Visual display requirements", Amendment 1: Annex C (normative): Visual performance and comfort test, (1992).

[13] Microsoft Research, "CHI 2013: An Immersive Event (Illusions create an Immersive Experience)." http://research.microsoft.com/en-us/news/features/chi2013-042913.aspx, (2013).

[14] Jones, Lynette A. "Kinesthetic Sensing." *Proceedings of Workshop on Human and Machine Haptics*, (2000): 1-10.

[15] Kieras, David. "Using the keystroke-level model to estimate execution times." *Technical Report* University of Michigan. (2001).

[16] Levin, Michael, and Woo, Alfred. "Tactile-feedback solutions for an enhanced user experience." *Information Display* 25.10 (2009): 18-21.

[17] MacKenzie, I. Scott. "Fitts' law as a research and design tool in human-computer interaction." *Human-computer interaction* 7.1 (1992): 91-139.

[18] MacKenzie, I. Scott. "Movement time prediction in human-computer interfaces." *Proceedings of Graphics Interface*. 92 (1992).

[19] Marois, Rene, and Ivanoff, Jason. "Capacity limits of information processing in the brain." *Trends in cognitive sciences* 9.6 (2005): 296-305.

[20] Miller, George A. "The magical number seven, plus or minus two: some limits on our capacity for processing information." *Psychological review* 63.2 (1956): 81.

[21] Ni, Tao, Schmidt, Greg S., Staadt, Oliver G., Livingston, Mark A., Ball, Robert, and May, Richard. "A survey of large high-resolution display technologies, techniques, and applications." *Virtual Reality Conference,* (2006).

[22] Norman, Donald A., and Stephen W. Draper. *User centered system design; new perspectives on human-computer interaction*. CRC Press, (1986).

[23] Nothdurft, Hans-Christoph. "The role of features in pre-attentive vision: comparison of orientation, motion and color cues." *Vision research* 33.14 (1993): 1937-1958.

[24] Olson, Harry Ferdinand. *Music, physics and engineering*. Dover Publications, (1967).

[25] Polk, T. A.; C. M. Seifert (2002). Cognitive Modeling. Cambridge, MA: MIT Press. ISBN 0-262-66116-0.

[26] Regenbrecht, Holger, Hauber, Joerg, Schoenfelder, Ralph, and Maegerlein Andreas. "Virtual reality aided assembly with directional vibro-tactile feedback." *Proceedings of the 3rd international conference on Computer graphics and interactive techniques in Australasia and South East Asia*. ACM, (2005): 381-387.

[27] Reeves, Leah M., et al. "Guidelines for multimodal user interface design." *Communications of the ACM* 47.1 (2004): 57-59.

[28] Salvucci, D. D., & Taatgen, N. A. (2008). Threaded cognition: An integrated theory of concurrent multitasking. Psychological Review', 130(1)', 101-130

[29] Sanders, Mark S., and McCormick, Ernest J. *Human factors in engineering and design*. McGraw-Hill Science, (1987).

[30] Schulz, Trenton. "Using the keystroke-level model to evaluate mobile phones." *Proceedings of IRIS31-The 31st Information Systems Research Seminar in Scandinavia*. (2008).

[31] Ware, C. (2012) Information Visualization, Third Edition: Perception for Design, Morgan Kaufmann.

[32] Weinstein, Sidney. "Intensive and extensive aspects of tactile sensitivity as a function of body part, sex and laterality." *the First Intl. Symp. on the Skin Senses,* (1968).

[33] Wolfe, Jeremy M., Treisman, Anne, and Horowitz, Todd S. "What shall we do with the pre-attentive processing stage: Use it or lose it?" *Journal of Vision* 3.9 (2003): 572-572.

[34] Apple “iPhone 4S” http://www.apple.com/ios/siri, (2013)

[35] Dougherty Patrick “Somatosensory Systems” http://neuroscience.uth.tmc.edu/s2/chapter02.html, (2013).

[36] IBM Research “Building Watson” http://www.whitehouse.gov/sites/default/files/ibm_watson.pdf, (2013)

[37] Patel Prachi "Synthetic Skin Sensitive to the Lightest Touch” http://spectrum.ieee.org/biomedical/bionics/synthetic-skin-sensitive-to-the-lightest-touch, (2013).

[38] Wikipedia “Haptics” http://en.wikipedia.org/wiki/Haptic, (2013).