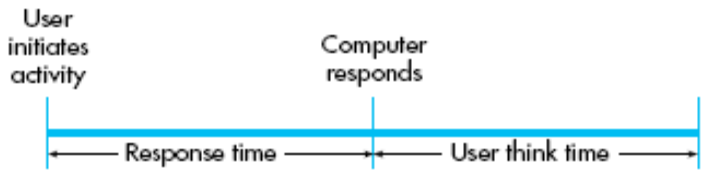
Quality of Service

Quality of Service

- **Introduction**
  - **1960s: mathematical computation association with computation time**
  - **World wide web: means graphics, & network congestion effect response time**
  - **Time is precious**
    - **Lengthy or unexpected system response time can produce:**
      - **Frustration**
      - **Annoyance**
      - **Eventual anger**
    - **Speedy and quickly done work can result in users:**
      - **learning less**
      - **reading with lower comprehension**
      - **making more ill-considered decisions**
      - **committing more data-entry errors**

Models of response-time impacts

- **Response time**
  - **The number of seconds it takes from the moment users initiate an activity until the computer presents results on the display**
- **User think time**
  - **The number of seconds the user thinks before entering the next action**



Models of response-time impacts

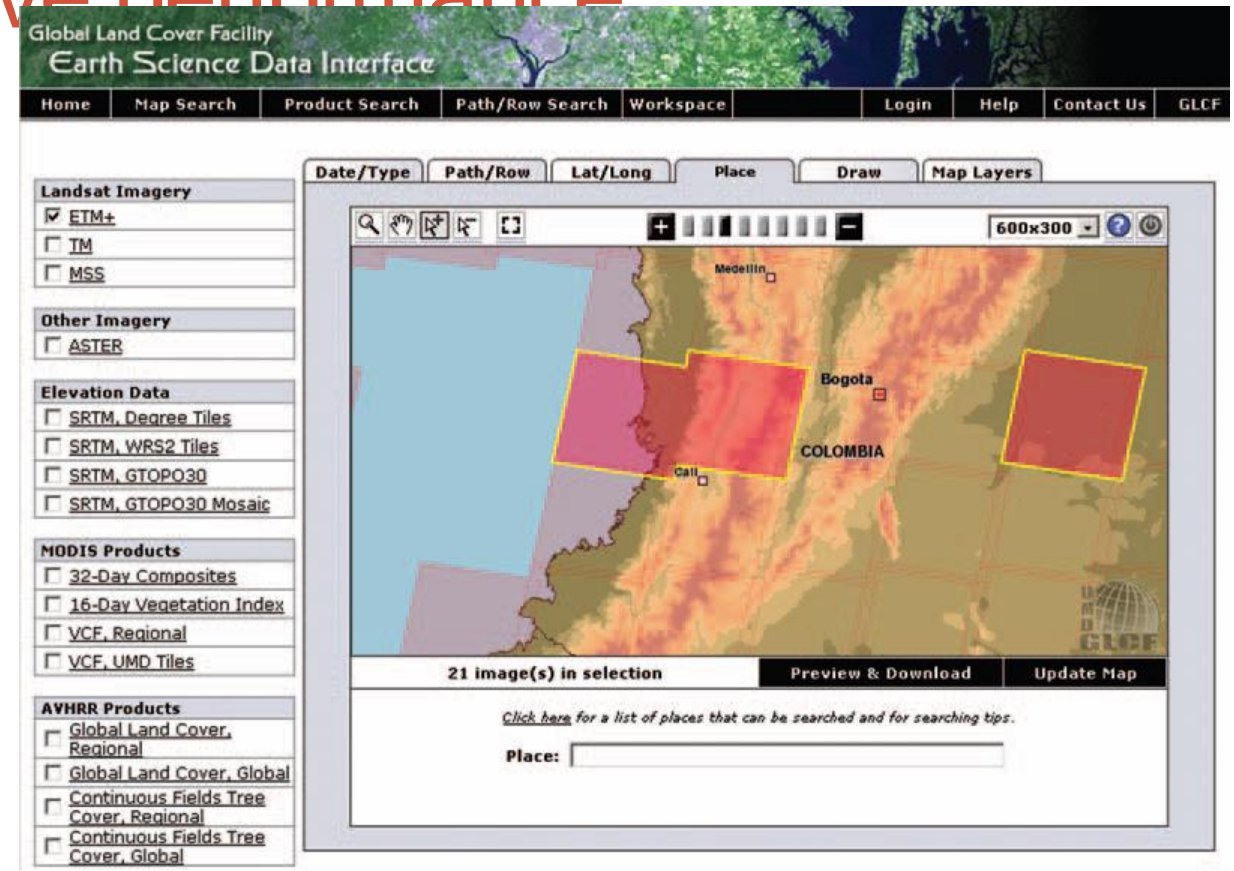
- **Designers of response times and display rates in HCI must consider:**
  - **complex interaction of technical feasibility**
  - **cost**
  - **task complexity**
  - **user expectations**
  - **speed of task performance**
  - **error rates**
  - **error handling procedures**

- **Overall majority of users prefer rapid interactions**
  - **Lengthy response times (15 seconds) are detrimental to productivity**
  - **Rapid response times (1 second or less) are preferable, but can increase errors for complex tasks**

Models of response-time impacts

- **Display Rate**
  - **Alphanumeric displays: The speed in characters per second at which characters appear for the user to read**
  - **World Wide Web Applications: Display rate may be limited by network transmission speed or server performance**
- **Reading textual information from a screen is a challenging cognitive and perceptual task**
  - **Users relax when the screen fills instantly- beyond a speed where someone may feel compelled to keep up**
- **Cognitive human performance would be useful for:**
  - **making predictions**
  - **designing systems**
  - **formulating management policies**

Tune World Wide Web applications to improve performance



Designers can optimize web pages to reduce byte counts and numbers of files or provide previews of materials available in digital libraries or archives to help reduce the number of queries and accesses to the network

Models of response-time impacts

Limitations of short-term and working memory

- **Any cognitive model must emerge from an understanding of human problem-solving abilities**
- **Magic number seven - plus or minus two**
  - **The average person can rapidly recognize seven chunks of information at a time**
  - **This information can be held for 15 to 30 seconds in short-term memory**
  - **Size of the chunks depends on the person' s familiarity with the material**
- **Short-term memory and working memory are used in conjunction for processing information and problem solving**
  - **Short-term memory processes perceptual input**
  - **Working memory generates and implements solutions**
- **People learn to cope with complex problems by developing higher-level concepts using several lower-level concepts brought together into a single chunk**
- **Short term and working memory are highly volatile**
  - **Disruptions cause loss of memory**
  - **Delays require that memory be refreshed**

Models of response-time impacts

Source of errors

- **Solutions to problems must be recorded to memory or implemented**
  - **Chance of error increases when solutions are recorded**
- **When using an interactive computer system users may formulate plans and have to wait for execution time of each step**
- **Long (1976) found unskilled and skilled typists worked more slowly and made more errors with longer response times**
  - **For a given user and task, there is a preferred response time**

Conditions for optimum problem solving

- **Longer response time causes uneasiness in the user because the penalty for error increases**
- **Shorter response time may cause the user to fail to comprehend the presented materials**
- **Progress indicators shorten perceived elapsed time and heighten satisfaction:**
  - **graphical indicators**
  - **blinking messages**
  - **numeric seconds left for completion**

Models of response-time impacts

Conditions for optimum problem solving (cont.)

- **Rapid task performance, low error rates, and high satisfaction can come from:**
  - **Users have adequate knowledge of the objects and actions necessary for the problem-solving task**
  - **The solution plan can be carries out without delays**
  - **Distractions are eliminated**
  - **User anxiety is low**
  - **There is feedback about progress toward solution**
  - **Errors can be avoided or handled easily**
- **Other conjectures in choosing the optimum interaction speed**
  - **Novices may exhibit better performance with slower response time**
  - **Novices prefer to work at slower speeds**
  - **With little penalty for an error, users prefer to work more quickly**
  - **When the task is familiar and easily comprehended, users prefer more rapid action**
  - **If users have experienced rapid performance previously, they will expect in future situations**

Expectations and attitudes

- **Related design issues may clarify the question of acceptable response time**
  - **E.g. how long before hearing a dial-tone**
- **Two-second limit (Miller, 1968) appropriate for many tasks**
- **But users have adapted a working style and expectation based on responses within a fraction of a second**
- **People can detect 8% changes in a 2-4 second response time**

Response-time choke

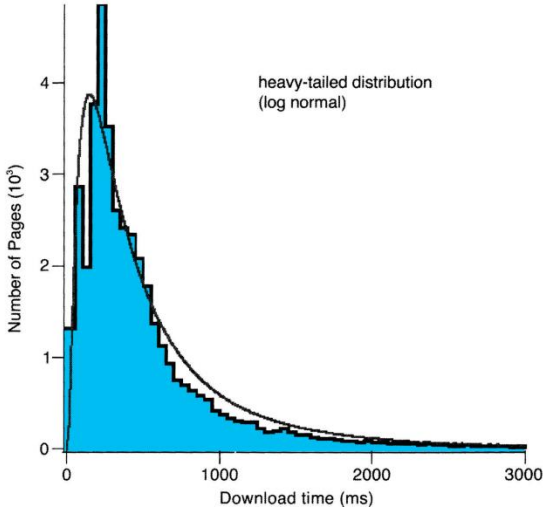
- **A system is slowed down when the load is light and potential performance high**
- **Makes the response time more uniform over time and across users, avoiding expectations that can't always be met**

Response time across web sites varies

It effects user interest and quality assessment

Three things influence response-time:

- **Previous experiences**
- **The individual's tolerance for delays**
- **Task complexity**

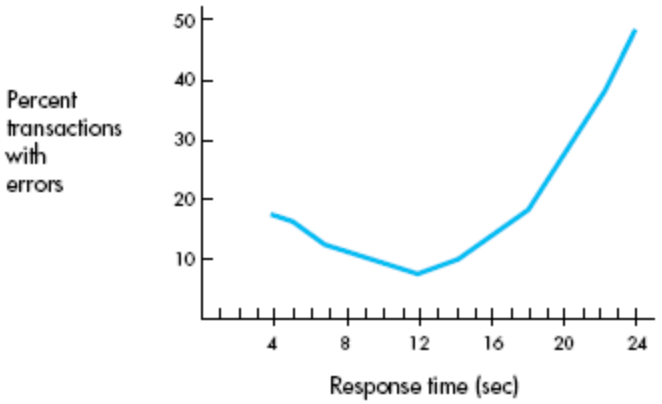


User productivity

- **Repetitive tasks**
  - **Nature of the task has a strong influence on whether changes in response time alter user productivity**
  - **Shorter response time means users responds more quickly, but decisions may not be optimal**
  - **Goodman and Spence (1981) – reduced response time lead to more productivity**
  - **Teal and Rudnecky (1992) – slower response time lead to more accuracy**
- **Problem solving tasks**
  - **Users will adapt their work style to the response time**
  - **Users will change their work habits as the response time changes**
  - **Grossberg, Wiesen, and Yntema (1976) – the time to solution was invariant with respect to response time**
- **Summary**
  - **Users pick up the pace of the system to work more quickly with shorter response time**
  - **Higher throughput of work demands more attention must be paid to minimizing the cost of delay of error recovery**

Variability

- **People are willing to pay substantial amounts of money to reduce the variability in their life, e.g. insurance**
- **Goodman and Spence (1981)**
  - **Subjects took more advantage of fast response time by making their subsequent commands immediately and balancing the time lost in waiting for slower responses**
- **Modest variations in response time (plus or minus 50% of the mean) appear to be tolerable**
- **It may be useful to slow down unexpected fast responses**



Frustrating experiences

- (Ceaparu et al., 2004) 46% to 53% of users' time was seen as being wasted
- Recommendations include improving the quality of service and changes by the user
- Poor quality of service is more difficult in emerging markets and developing nations
- User training can help
- Email a common application, but also a common source of frustration
- Viruses also a problem

Frustrating experiences (cont.)

- Since frustration, distractions, and interruptions can impede smooth progress, design strategies should enable users to maintain concentration.
- Three initial strategies can reduce user frustration:
  1. Reduce short-term and working memory load
  2. Provide information abundant interfaces
  3. Increase automaticity
     - Automaticity in this context is the processing of information (in response to stimuli) in a way that is automatic and involuntary, occurring without conscious control.
     - An example is when a user performs a complex sequence of actions with only a light cognitive load, like a driver following a familiar route to work with little apparent effort.