QUALITY OF SERVICE
UNIT 4

Dr. P.V. S LAKSHMI JAGADAMBA
Professor, Department of CSE,
GVP College of Engineering for Women

INTRODUCTION

- 1960s: mathematical computation association with computation time
- Time-shared systems: more reasons for delays: contention for computational resources
- World wide web: means graphics, & network congestion effect response time

- All these concerns are usually discussed under the term *Quality of Service (QoS)*

BASIC HUMAN VALUES

1. Time is precious
   - Lengthy or unexpected system response time can produce:
     - Frustration
     - Annoyance
     - Eventual anger

- *which lead to frequent errors and low satisfaction*

► BASIC HUMAN VALUES

2. “Harmful mistakes should be avoided”

- This may sometimes means the pace of work must slow.
- 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

- *Stress can build in all these situations, especially if the damage is big.*

BASIC HUMAN VALUES

3. “Reduce user frustration”

- Frustration results in making mistakes and giving up working
- Causes of frustration:
  - Long delays
  - Crashes that destroy data
  - Software bugs that produce incorrect results
  - Poor design that lead to user confusion
- Network environments generate further frustrations:
  - Unreliable service providers
  - Dropped lines
  - Email spam, and viruses

QUALITY OF SERVICE

- Definition: QoS = a set of concerns that deal with issues such as reliability or performance (e.g., response times, server failure, etc.)
- Why is it important in the context of HCI?
  - Influences the task performance/progress on a task
  - Influences the error rate
  - Influences subjective satisfaction

QUALITY OF SERVICE

- Responsibility of the UI designer:
  - Design UIs in a way that high (or rather well balanced) QoS is guaranteed
  - Avoid harmful mistakes, save time, reduce frustrations
  - Balance rapid performance with low error rates
- For this purpose the UI designer must understand certain QoS aspects and how they affect the user/task performance

QUALITY OF SERVICE

- Quality of service is mostly effected by decisions made by
  - Network designers and operators
  - Interface designers and builders
    - reduce byte count for web pages
    - reduce number of queries and access to the network
    - Users may have the opportunity to choose from fast or slow services and from viewing low-resolution versus high-resolution images

- For users the main concern for quality of service is computer *response time*.

QUALITY OF SERVICE ISSUES

- Issues of Response time
  1. Model of Response-Time Impacts
  2. User Expectations and Attitudes
  3. User Productivity
  4. Variability in Response Times
  5. Frustrating Experiences
- Other issues
  - Application crashes
  - Unreliable network services
  - Malicious threats

QUALITY OF SERVICE

ISSUES OF RESPONSE TIME

1. MODEL OF RESPONSE-TIME IMPACTS

2. USER EXPECTATIONS AND ATTITUDES

3. USER PRODUCTIVITY

4. VARIABILITY IN RESPONSE TIMES

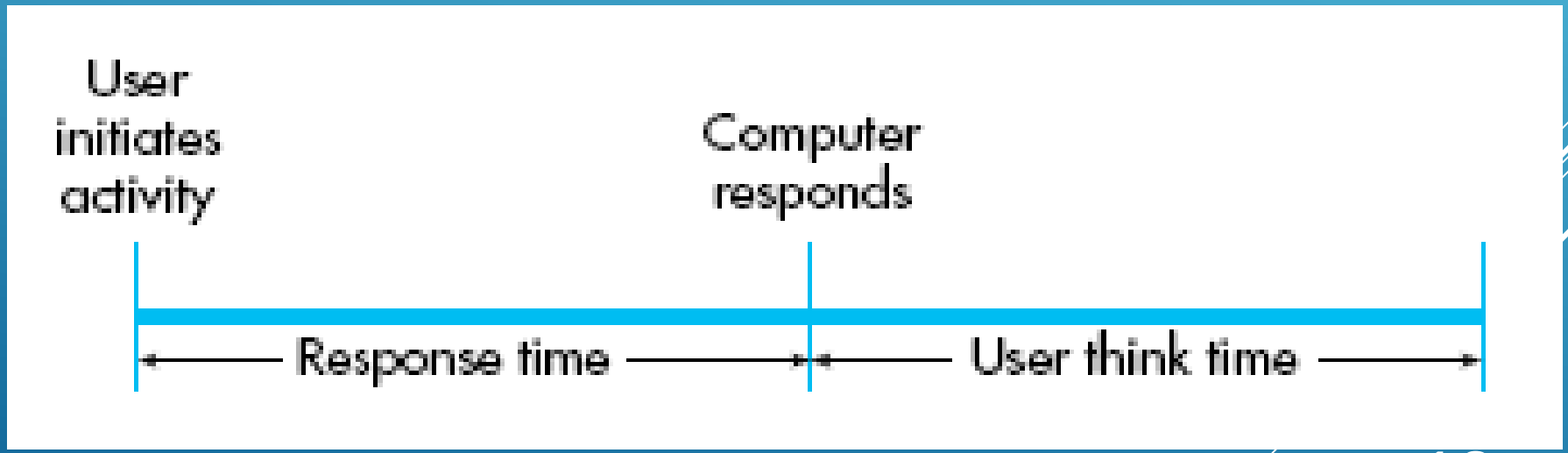
5. FRUSTRATING EXPERIENCES

MODEL OF RESPONSE-TIME IMPACTS

- Response Time = Number of seconds it takes from the moment a user initiates an action (e.g., by pressing a button or ENTER) until the computer begins to present results on the display, printer, etc.
  - The response commonly leads to the user formulating a new goal
  - Computer response times can be measured and/or predicted
- User Think Time = Time that users think before initiating the next action
  - User think times are very difficult to measure or predict

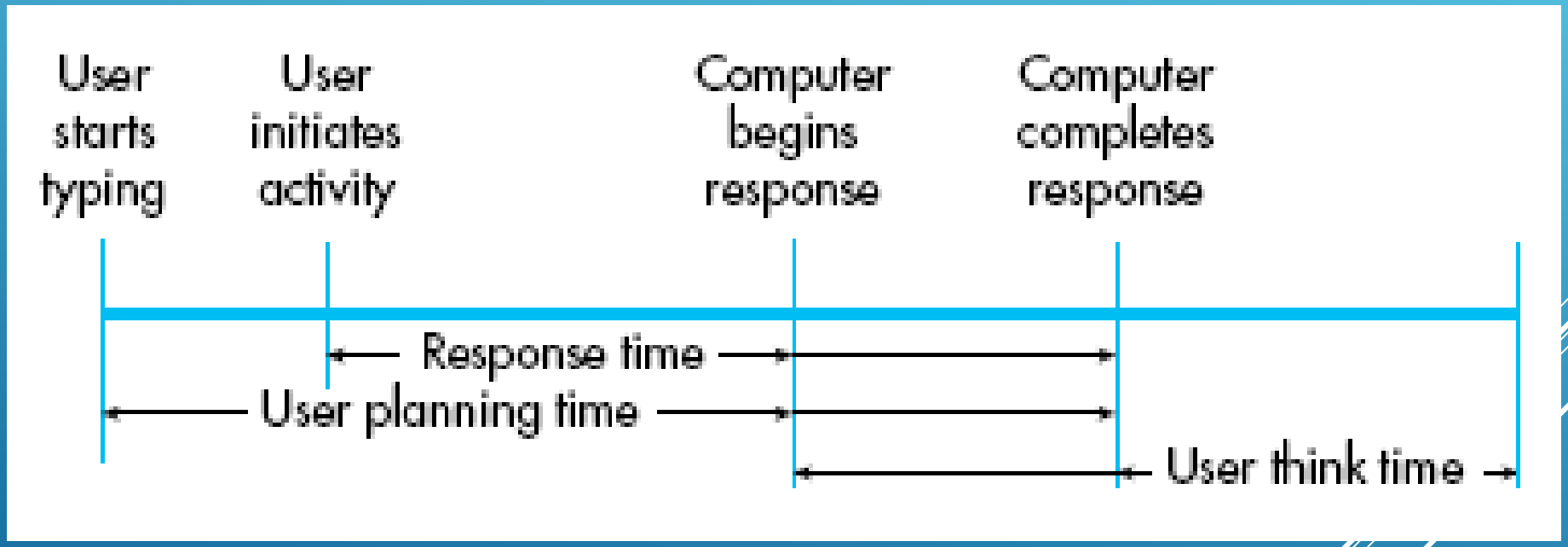
MODEL OF RESPONSE-TIME IMPACTS

- Simple model of response time
  - Users (1) initiate, (2) wait for response, (3) watch results, (4) think for a while, and initiate again
    - Response time (?)
    - Think time (?)



MODEL OF RESPONSE-TIME IMPACTS

- More realistic model of response time
  - People will use whatever time they have to plan ahead



MODEL OF RESPONSE-TIME IMPACTS

- For a given user and task, there is a preferred response time

Long response time	Short response time
Lead to wasted effort and more errors, because the solution plan must be reviewed repeatedly	May generate a faster pace in which solution plans are prepared hastily and incompletely
Causes uneasiness because the penalty for error increases	The user may pick up the pace of interface and fail to fully comprehend the presented materials

MODEL OF RESPONSE-TIME IMPACTS

- Overall majority of users prefer rapid interactions, however, overall *productivity* depends on
  - interaction speed
  - error rates
  - ease of recovery from errors
- Lengthy response times (>15 seconds) are harmful to productivity
  - increasing error rates and decreasing satisfaction
- Rapid response times (1 second or less) are preferable, but can increase errors for complex tasks if the user does not spent sufficient time to think.
- The high cost of providing rapid response times and the loss from increased errors must be evaluated in the choice of an optimum pace

MODEL OF RESPONSE-TIME IMPACTS

- Display Rate
  - Alphanumeric displays: The speed in characters per second at which characters appear for the user to read. e.g., 120cps for mobile devices
  - World Wide Web Applications: Display rate may be limited by network transmission speed or server performance

MODEL OF RESPONSE-TIME IMPACTS

Challenges

- General issues that need to be considered by designers and/or network managers when specifying QoS levels:
  - Technical feasibility
  - Costs
  - Task complexity
  - User expectations
  - Speed of performance
  - Error rates & Error-handling procedures
- This is furthermore influences by:
  - Different user personalities
  - Familiarity with computers
  - Different experiences with tasks
  - Motivation

MODEL OF RESPONSE-TIME IMPACTS

Cognitive Models of Human Performance

- Can help to determine the *User Think Time* and based on it the appropriate *Computer Response Time*
- Problem: Such models are very hard to create
- Example:
  - Look at the short and long term memory
  - How do both work, how many chunks of information can users comprehend and process per second?
  - Do users plan ahead in there short term memory?

✓ Answering such questions might help to determine how long users need to compile/calculate a task/action before they initiate the next action but also how long a response can be before the task productivity goes down

MODEL OF RESPONSE-TIME IMPACTS

- Reading textual information from a screen is a challenging cognitive task
  - Users relax when the screen fills instantly
  - It is useful to display text first, leaving space for the graphical elements
- Short term and working memory are highly volatile
  - Disruptions cause loss of memory
  - Delays require that memory be refreshed
  - Visual distractions, noisy environments, and anxiety interfere with cognitive processing

MODEL OF RESPONSE-TIME IMPACTS

- When using an interactive computer system users may formulate plans and have to wait for execution time of each step
- If there is an unexpected result (error), or long delay, then users may forget part of the plan or be forced to review the plan continually

MODEL OF RESPONSE-TIME IMPACTS

Limitations of short-term and working memory

- Magic number
  - 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*

MODEL OF RESPONSE-TIME IMPACTS

Ideal Conditions

- Users have adequate knowledge of objects and actions necessary for the problem-solving task
- Solution plan can be carried out without delays
- Distractions are eliminated
- User anxiety is low
- There is feedback about progress towards the solution
- Errors can be avoided or, if they occur, can be handled easily

✓ These ideal conditions help to achieve rapid task performance, low error rates and high satisfaction.

MODEL OF RESPONSE-TIME IMPACTS

Other Issues

- Novices may exhibit better performance with somewhat slower response times
- Novices prefer to work at speeds slower than those chosen by knowledgeable, frequent users
- When there is 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 and demand it in the future

QUALITY OF SERVICE

ISSUES OF RESPONSE TIME

1. MODEL OF RESPONSE-TIME IMPACTS

2. USER EXPECTATIONS AND ATTITUDES

3. USER PRODUCTIVITY

4. VARIABILITY IN RESPONSE TIMES

5. FRUSTRATING EXPERIENCES

USER EXPECTATIONS AND ATTITUDES

- Overview Questions:
  - How long will users wait for a computer to respond before they will become frustrated?
  - Will users be more happily to wait for a valued document than a low-quality output?

- Related design issues may clarify the question of acceptable response time
  - Two-second limit appropriate for many tasks
  - But users have adapted a working style and expectation based on responses within a fraction of a second. e.g., key typed, wheel turn, …
  - In other situations, users are accustomed to longer response times. e.g., traffic light

USER EXPECTATIONS AND ATTITUDES

- It is important that designer need to understand such issues

- Factors that influence acceptable response time:
  1. Prior established expectations
  2. Users' individual tolerance for delays
  3. Task difference

USER EXPECTATIONS AND ATTITUDES

1. Prior established expectations

- Based on their past experiences, users often have established expectations in regard to the time that is required to complete a task
  - What would be your reaction when the system response is:
    - Almost as you expected
    - Later than expected
    - Sooner than expected
    - Very much sooner than expected
  - 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

USER EXPECTATIONS AND ATTITUDES

1. Prior established expectations

- Consequence: Be careful with varying performance in different situation
  - Performance up/down may lead to concern/frustration
  - Need to balance concern and frustration
- Example: Network speed
  - "Typical" usage times (fast) vs. peak times (slow)
  - Purposely slowing the network down during "typical" usage times may have a positive effect on the user experience

USER EXPECTATIONS AND ATTITUDES

2. Users' individual tolerance for delays

- There is a large variation in what users consider being an acceptable waiting time
- Influencing factors:
  - Personality
  - Mood
  - Time of the day
  - Skill level (novice vs. expert) – Novice users may wait much longer
  - Perception of pressure to complete a task
  - Cost
  - Age
  - Cultural context

USER EXPECTATIONS AND ATTITUDES

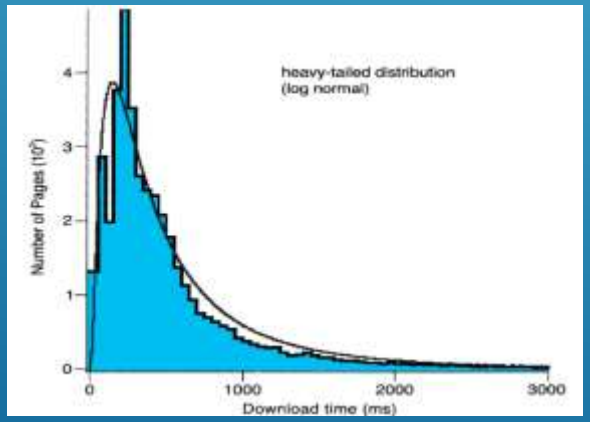
3. Task Difference

- Task Difference is task complexity and user familiarity with the task
- Task Difference:
  - Repetitive tasks that require little problem solving
    - Performance expectations rather high / low response times
  - Complex tasks with several action options at every step
    - Response time not as crucial: User plans ahead during longer response times
  - Real-time response tasks
    - Some tasks simply require a fast response (less than 10 ms)

USER EXPECTATIONS AND ATTITUDES

3. Task Difference

- Some tasks place high demands on rapid system performance
  - e.g., User-controlled 3D animations, simulators, VoIP telephony
- The range of response time is highly varied across web sites
  - As response times increase, users find web-page content less interesting and
  - lower in quality may affect a company's image.



USER EXPECTATIONS AND ATTITUDES

General observations:

In summary, three conjectures emerge:

1. Individual differences are large and users are adaptive. They will work faster as they gain experience and will change their working strategies as response time change. It may be useful to allow people to set their own pace of interaction (e.g., in games)
2. For repetitive task, user prefer and will work more rapidly with short response times.
3. For complex tasks, users can adapt to working with slow response time with no loss of productivity, but their dissatisfaction increases as response time lengthen.

Finally

- Users are highly adaptive and can change their working style to accommodate differences in response time (e.g., slow down, multitask, plan ahead)
- Still: Dissatisfaction grows with growing response times

QUALITY OF SERVICE

ISSUES OF RESPONSE TIME

1. MODEL OF RESPONSE-TIME IMPACTS

2. USER EXPECTATIONS AND ATTITUDES

3. USER PRODUCTIVITY

4. VARIABILITY IN RESPONSE TIMES

5. FRUSTRATING EXPERIENCES

USER PRODUCTIVITY

Overview

- Response times ↓ ⇒ productivity ↑
- Long response times may lead to users finding shortcuts or ways for concurrent processing
- Working to quickly may lead to errors and can decrease productivity

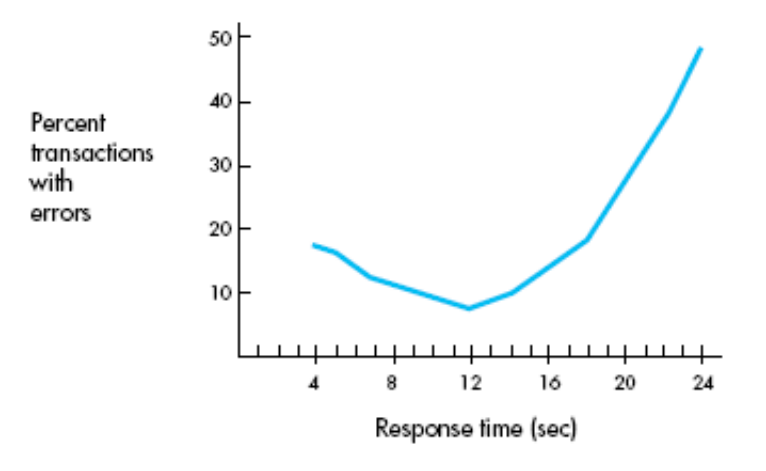
HCI goal: Increase productivity: Requires studying response times and user behavior.

Problem: The response time/productivity relation changes based on the task and user

1. Repetitive control task (may have more influence)
   - Shorter response time means users responds more quickly
     - decisions may not be optimal, but penalty for a poor choice is small
   - Reduced response time lead to more productivity (Goodman and Spence)
   - Slower response time lead to more accuracy (Teal and Rudnecky)

USER PRODUCTIVITY

2. Complex problem solving tasks actions (less influence)
   - Users will adapt their work style to the response time
   - The time to solution was invariant with respect to response time (According to Grossberg, Wiesen, and Yntema)
   - Error rates were lowest as 12 sec response time, but productivity increased linearly with reduction in response time (According to Barber and Lucas).



USER PRODUCTIVITY

Some studies and their results

- Control Task
  - R < 1 sec. leads to good productivity
- Data Entry Task
  - R < 1 sec. Leads to poor productivity (errors may occur through anticipation behavior)
- Complex Task
  - Invariant: Users will adapt
- Statistical Problem Solving Tasks
  - Invariant: Users will adapt

USER PRODUCTIVITY

Summary

- Users pick up the pace of the interface
- Users consistently prefer a faster pace
- Error rates at shorter response times increase with the cognitive complexity of the task
- Each task appears to have an optimal pace and response times that are shorter or longer than this pace lead to increased errors

QUALITY OF SERVICE

ISSUES OF RESPONSE TIME

1. MODEL OF RESPONSE-TIME IMPACTS

2. USER EXPECTATIONS AND ATTITUDES

3. USER PRODUCTIVITY

4. VARIABILITY IN RESPONSE TIMES

5. FRUSTRATING EXPERIENCES

VARIABILITY IN RESPONSE TIME

Overview

- Users get confused if they encounter response times that are significantly below or above the expected or typical response time
- If the variation is high (e.g., 0.5 or 15 sec. for a 3 sec. response) this may be detected
- As a UI designer one need to indicate such unusual fast or slow responses
- If the variation is more modest (e.g., 75% of participants in a study were able to detect 8% variations in 2-4 sec.)
- As a UI designer it may be advisable to put restrictions on response time variability
- Problem: Technical feasibility
  - Led to specifying response time for action classes (e.g., 1, 2, 4)

VARIABILITY IN RESPONSE TIME

Studies suggest (summary)

- Modest variations in response time (+/- 50% of the mean) appear to tolerable and have little effect on performance
- Frustration emerges only if delays are unusually long (≥ twice the anticipated time)
- Anxiety about an erroneous command may emerge only if the response time is unusually short (roughly 25% of the anticipated time)
- Even with extreme changes users appear to be adaptable enough to complete their task

VARIABILITY IN RESPONSE TIME

Implications on Web Sites

- Guidelines (examples)
  - Look at the average download rate and adjust your byte count
  - Load text first
  - Give users control over size/quality of multimedia elements (images/video/sound)
- Other issues
  - Tradeoff: Response time vs. resource expenditure
  - Studies suggest that high response times lead to:
    - Reduced interest in a web-site
    - Reduced quality perception of sites
    - Company images ↓

QUALITY OF SERVICE

ISSUES OF RESPONSE TIME

1. MODEL OF RESPONSE-TIME IMPACTS

2. USER EXPECTATIONS AND ATTITUDES

3. USER PRODUCTIVITY

4. VARIABILITY IN RESPONSE TIMES

5. FRUSTRATING EXPERIENCES

FRUSTRATING EXPERIENCES

- 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.

ANY QUESTIONS???