

Human-Machine Interaction
Interaction Modalities &
Technologies

Visual and Auditory Interaction

Dr. Patrick Chan
patrickchan@scut.edu.cn

South China University of Technology

1

HMI Visual Interaction Foundations

Human Visual Dominance

- Rely heavily on vision for environmental awareness
 - Afraid in dark environments
 - Sleeping at night in ancient times
- Vision Characteristics:
 - High detail and sharpness
 - Strong ability to recognize patterns
 - Quick detection of unusual changes



2

Visual Information Characteristic

• High Bandwidth

- Simultaneous processing of **multiple elements**
- Parallel perception of shape, color, motion, depth

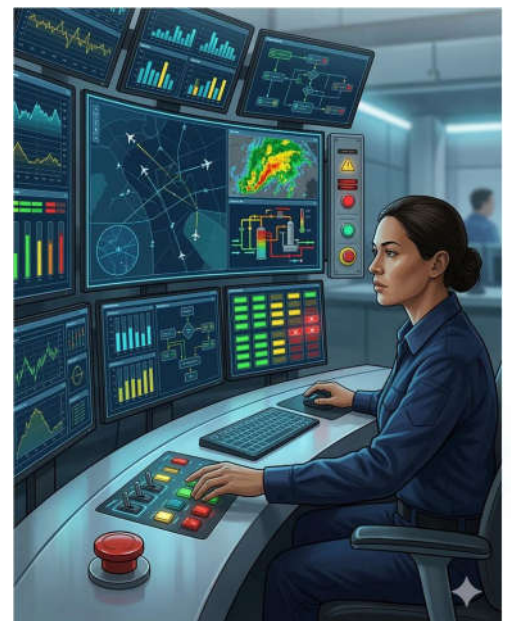
• Persistence

- Unlike audio, visual information can remain stable
- Information **remains visible**
- Supports comparison over time
- Allows **scanning** and re-checking

3

Vision in Safety-Critical Systems

- Visual displays are **commonly used in Safety-Critical Systems**, e.g. aviation, nuclear plants, medical systems, and autonomous vehicles
 - **Monitor** system state
 - **Detect** abnormal conditions
 - **Anticipate** failures
 - **Execute** corrective actions



4

Poor Visual Design

- **Cognitive load increases**
 - Slower reaction time
 - Mode confusion
 - Missed alarms
 - Loss of trust
- **Not usability** issues **but** also escalate into **safety failures**



5

Communication Channel

- **Visual interfaces** are **communication channels** between the system and the human
 - **Continuously communicates** through visual signals
 - Current system **state**
 - **Mode** of operation
 - **Progress** of tasks
 - System **confidence** in decisions



6

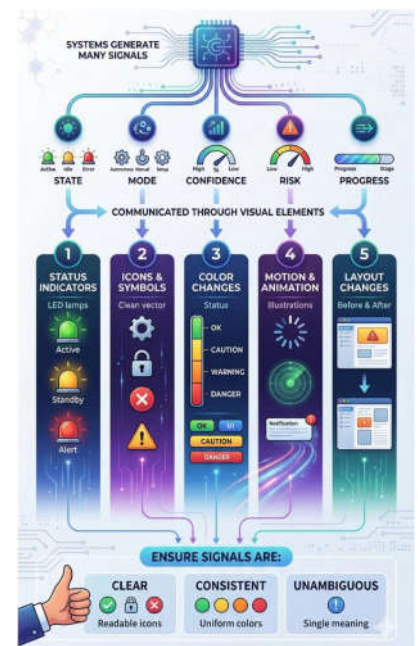
Communication Channel

- It is a **language spoken by the system**
 - Not decoration
 - Let's user to Interpret system states and detect change
- Examples
 - Smartphone: Loading spinner → system is processing
 - Autonomous vehicle: Blue light → autonomous mode active
 - Medical monitor: Red alarm → abnormal patient status



System-to-Human Signaling

- **Systems generate many signals** showing different state, mode, confidence, risk, progress
- Communicated through visual elements
 - status indicators
 - icons and symbols
 - color changes
 - motion and animation
 - layout changes



Visual Cues

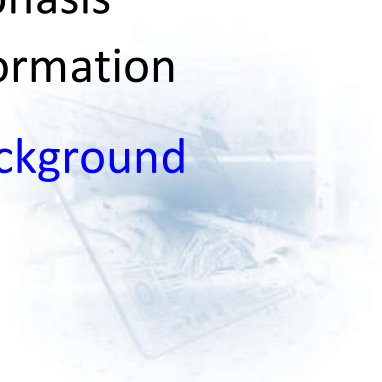
- **Explicit Cues**

- Text warnings
- Highlighted alerts
- Clear labels and messages
- Suitable for critical events

- **Implicit Cues**

- Subtle color variation
- Animation speed
- Spatial emphasis
- Icon transformation
- Support background awareness

- **Balance** prevents overload or invisibility



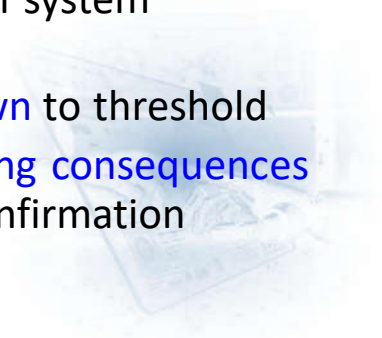
Feedback vs Feedforward

- **Feedback**

- Communicates what has already happened
- Examples:
 - Button changes color after activation
 - Status updates to “Completed”

- **Feedforward**

- Communicates what will happen next
- Examples:
 - Preview of system trajectory
 - Countdown to threshold
 - Highlighting consequences before confirmation



Temporal Behavior

- Visual information **is shown over time**
- Timing is critical and part of interface design
 - **Too slow** → perceived unresponsiveness
 - **Too fast** → missed signals
- Should show **intermediate feedback**
 - **< 100 ms** → feels instantaneous
 - **> 3 seconds** → user attention drops



UI Design

- Must support **safe, efficient, and understandable interaction, not** only about appearance
- Example
 - Aircraft cockpit displays prioritize **readability and function over decoration**



UI Design Goal

- **Clarity and predictability**
 - What the system is doing and what will happen next
- **Minimizing cognitive load**
 - Reduce mental effort and avoid unnecessary complexity
- **Supporting fast and correct decisions**
 - Easy to interpret under time pressure



13

UI Design Monitoring vs Control

- **Two roles** in HMI should be supported:
 - **Monitoring**
 - **Observing** system behavior
 - **Detecting** anomalies
 - **Maintaining** situational awareness
 - **Control**
 - **Issuing commands**
 - **Adjusting** parameters
 - **Intervening** in system behavior



14

State Visibility

- The operator must always understand:
 - **What** the system is doing
 - **Why** it is doing it
 - **What** it will do next
- **Situational awareness** depends on clear state visibility
 - If the state is unclear, safe cooperation is impossible

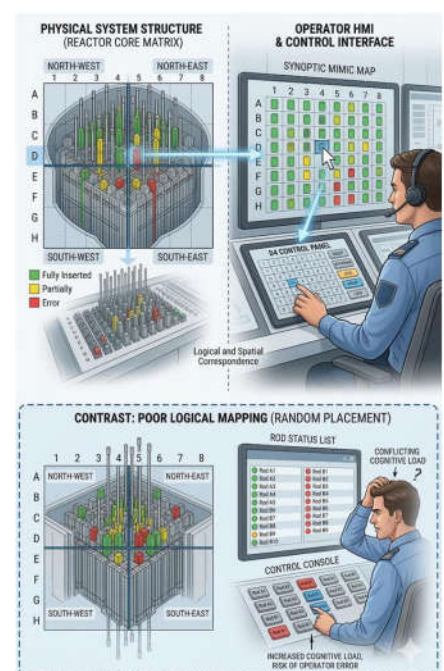


15

UI Design

Mapping UI & System Functions

- **Controls and displays** should have **logical relationships** with system functions.
 - **Avoid** user confusion and errors
 - Control location **reflects physical** system location
 - Interface structure **mirrors system** structure



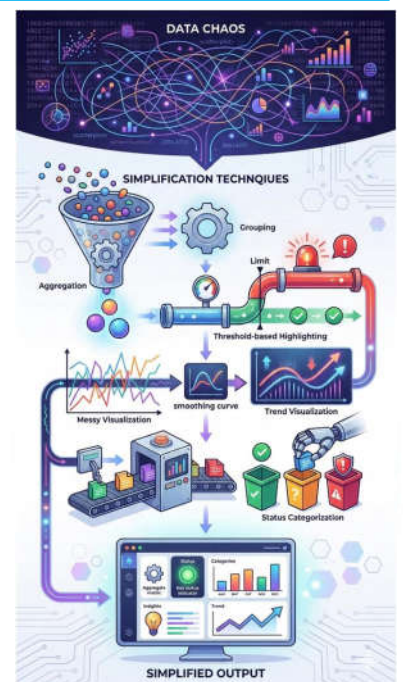
Mapping UI & System Functions

- Example: Car dashboard layout
 - **Left** → vehicle status
 - **Center** → navigation and media
 - **Right** → driver assistance



Meaningful Representation

- **Systems**
 - Detect numerical values
 - Process raw signals
- **Humans**
 - Interpret patterns
 - Rely on visual abstraction
- The interface must translate internal states into meaningful visual forms



Meaningful Representation

- **Raw data**

- High dimensional
- Continuous
- Numerical
- Machine-readable

- **Meaningful display**

- Simplified
- Structured
- Contextualized
- Human-interpretable

- **Core Question**

What does the operator actually need to see?



Abstraction

- **Reduces complexity without losing critical meaning**
- **Goal: Rapid understanding, not detailed computation**
- Techniques:
 - Aggregation
 - Threshold-based highlighting
 - Trend visualization
 - Status categorization



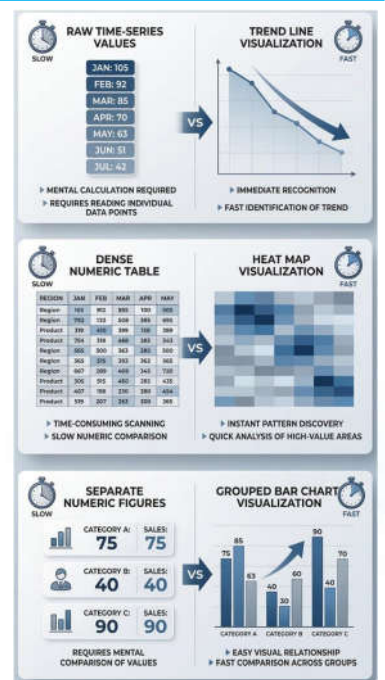
Mapping

- **Algorithm outputs** may include:
 - Probability scores
 - Confidence levels
 - Classification labels
 - Trajectory predictions
- These must be **mapped into**:
 - Visual cues
 - Status categories
 - Risk levels
 - Actionable insights
- The mapping must **align with human mental models**.



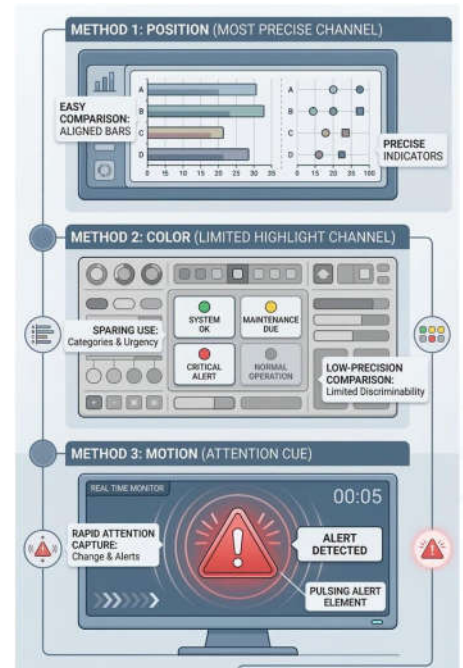
Pattern Recognition vs Numbers

- Humans **recognize patterns faster** than they **read numbers**
- Better examples:
 - Trend line instead of raw values
 - Heat map instead of table
 - Bar comparison instead of separate figures
- Visualization **reduces mental calculation**



Visual Encoding

- Information can be encoded using:
 - **Position:** Most precise visual channel, Good for comparisons
 - **Color:** Highlights categories or urgency, Use sparingly
 - **Motion:** Attracts attention, Suitable for change or alerts
- Encoding choice affects interpretation speed.



Temporal Consistency

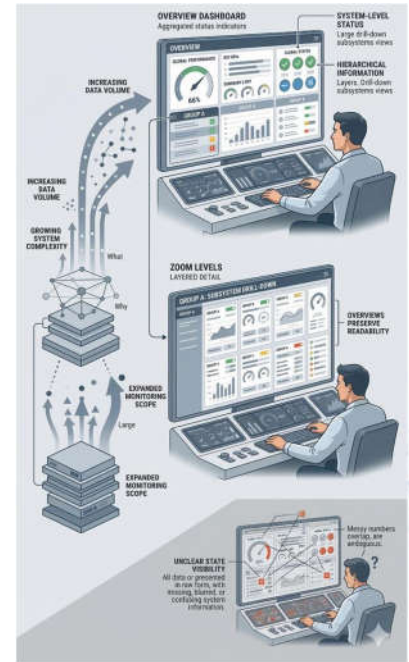
- Visual elements should **remain stable over time**.
- Avoid:
 - **Frequent** layout **shifts**
 - **Changing** color meanings
 - **Inconsistent** scales
- Temporal consistency **supports memory**



UI Design: Meaningful Representation

Scalability

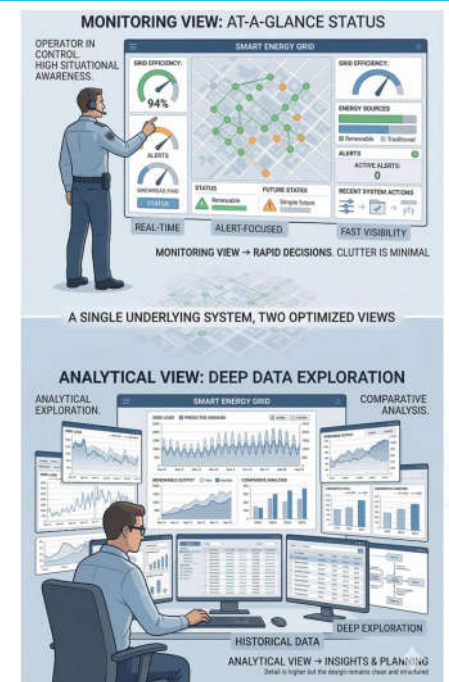
- Clarity must **scale with complexity**.
- Visualization must remain clear as:
 - Data volume increases
 - System complexity grows
 - Monitoring scope expands
- Design for:
 - Aggregation
 - Zoom levels
 - Layered detail



UI Design: Meaningful Representation

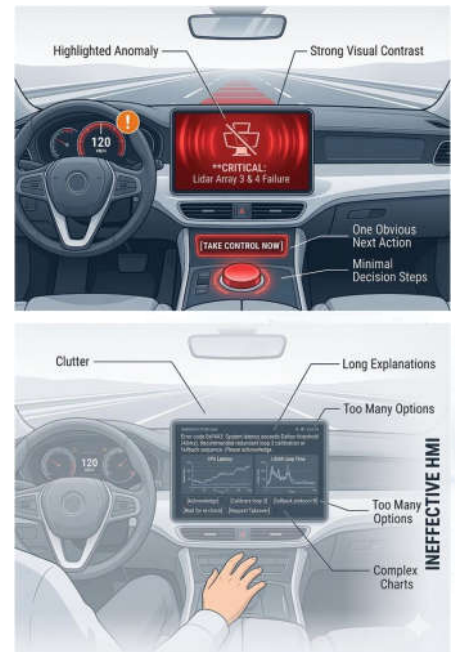
Different Views

- Different tasks require different visual structures
 - **Monitoring View**
 - Real-time status
 - Simplified
 - Alert-focused
 - **Analytical View**
 - Detailed exploration
 - Historical data
 - Comparative analysis



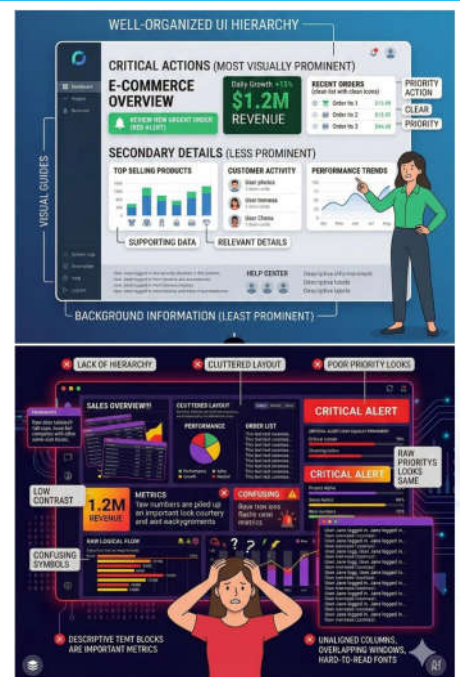
Time Pressure

- In **urgent** situations, **speed** and **clarity** become critical
- Users cannot:
 - Read long explanations
 - Interpret complex charts
- Design must:
 - Highlight anomalies clearly
 - Use strong visual contrast
 - Reduce decision steps



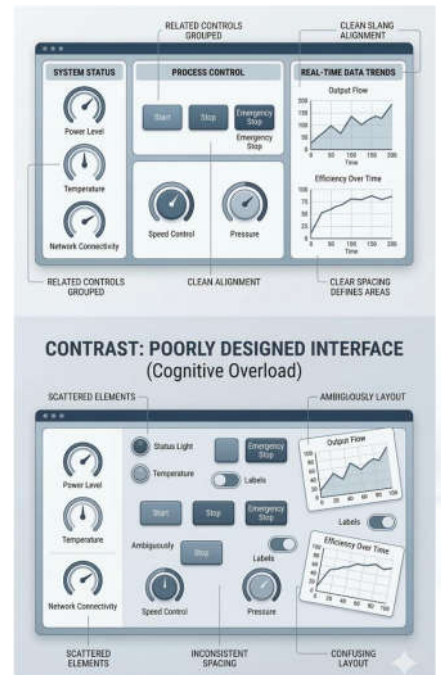
Hierarchy

- Information should be **organized in a hierarchy** to reflect its importance
 - Users should be able to **immediately distinguish the importance of information**
- Hierarchy is created using Size, Color, Position, Contrast, Motion



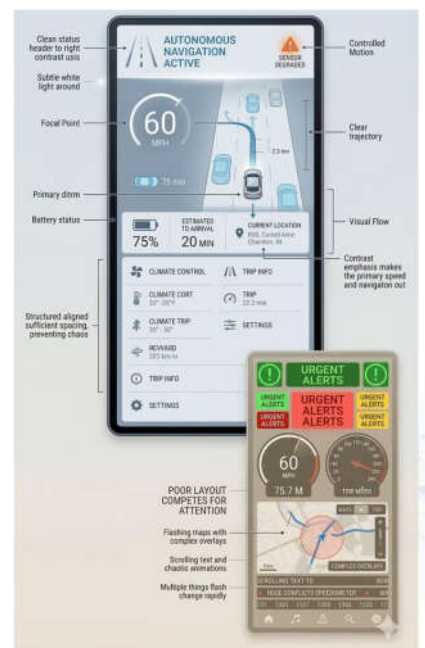
Grouping, Alignment & Spacing

- Human perception naturally **groups related elements**.
- Good UI design uses
 - **Grouping**: indicate related functions
 - **Alignment**: create structure
 - **Spacing**: separate different information areas



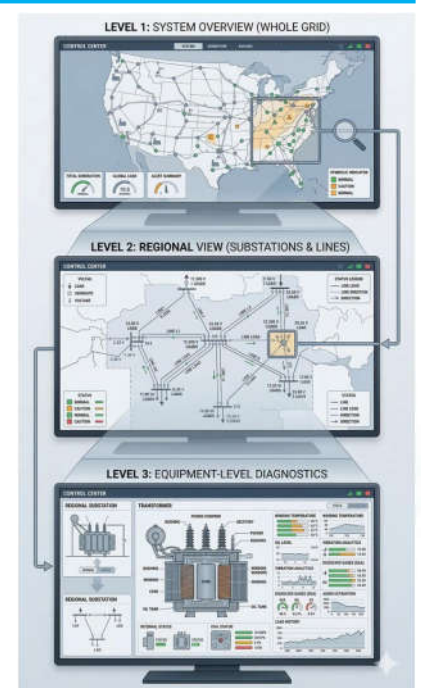
Attention Guidance

- **Guides** the user's **eyes naturally**
 - **Guide** attention, **not compete** for it
- Techniques:
 - **Visual flow** (top-to-bottom or left-to-right)
 - **Contrast** emphasis
 - Clear **focal** points
 - Controlled use of **motion**



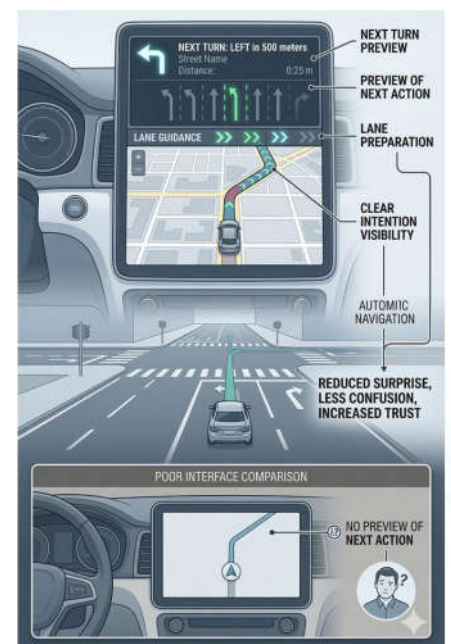
Progressive Disclosure

- **Not all information** should be **shown at once**
 - Show **essential** information **first**
 - Reveal more **detail when needed**
- Benefits
 - **Reduces cognitive load**
 - Keeps **interface clean**



Intention

- Must **inform** users that **what** the system is **trying to do**
 - **Reduces surprise and confusion**
 - Build **trust**
- Common methods
 - Visual **previews** before execution
 - Animated **indicators** showing **next action**



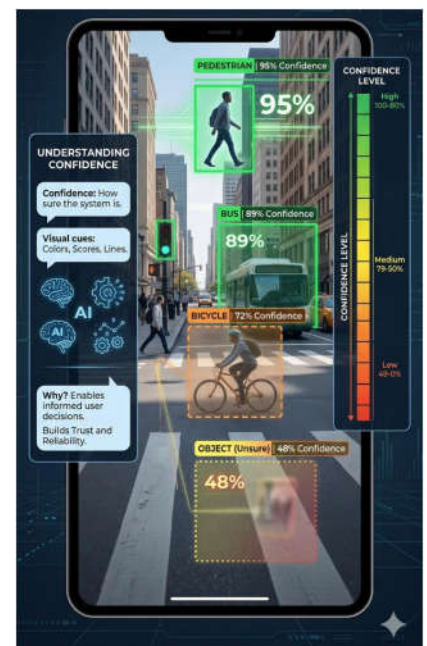
Transparency

- Intelligent systems are often **dynamic** and **unpredictable**
- Non-ideal system conditions should be clearly reflected
- Transparency increases the trust



Confidence

- **How certain** the system is
 - AI outputs may be uncertain
- Good design should
 - **Confidence score** should be indicated, especially when it is low
 - E.g. Object detection box with confidence
 - **Confirmation** should be **required** in **uncertain** cases
 - E.g. "Did you mean: Call Alex?"



Latency

- Intelligent systems **introduce delays**
 - Sensor latency
 - Network communication
 - AI model processing
 - Cloud inference
- Good design should show
 - Show **processing** status
 - System **mode**



Partial or Degraded Information

- Sometimes the received data is **incomplete or degraded**
 - Camera obstruction
 - Sensor failure
 - Poor environmental **conditions**
- Good design should show
 - Data **limitations**
 - **Reduced reliability**
 - System **fallback** behavior



UI Design: Meaningful Representation: Transparency

Information Overload

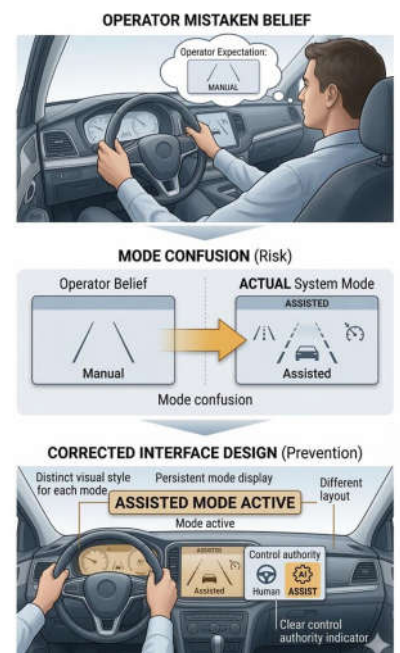
- **Transparency** usually provide **more information**
- More information is always better?
 - **Too little information**
→ users **cannot trust** the system
 - **Too much information**
→ users **cannot process** it
- Good design provides **the right information at the right time**



UI Design

Mode Transition

- Modern intelligent systems operate in **multiple modes**
 - Users have **different responsibility and expectation** in modes
- **Mode Confusion** should be **avoided**
 - **Persistent mode display**
 - **Distinct visual styles for each mode**
 - **Clear control authority indicators**



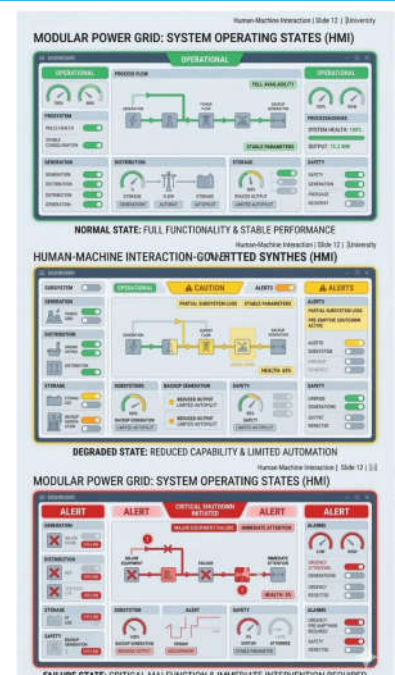
Transition Design

- Transitions must be **explicit**, **noticeable** and **understandable**
 - Clear visual indication
 - Change in color or layout
 - Status confirmation



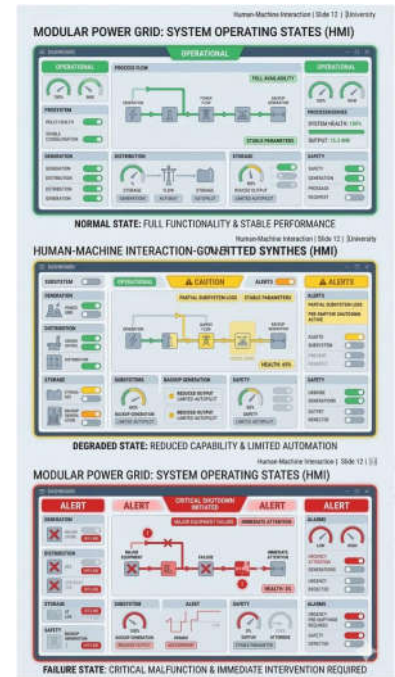
Normal, Degraded, and Failure

- Each **state** must be **clearly distinguishable**
- Systems may operate in:
 - **Normal State**: Full functionality, Stable performance
 - **Degraded State**: Reduced capability, Limited automation
 - **Failure State**: Critical malfunction, Immediate intervention required



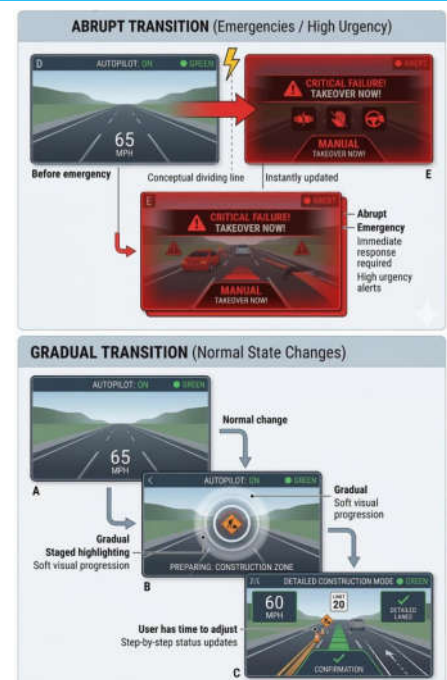
Visual Continuity

- Helps users **maintain mental models**
- Transitions should preserve:
 - **Layout** stability
 - Familiar **structure**
 - Consistent **encoding**



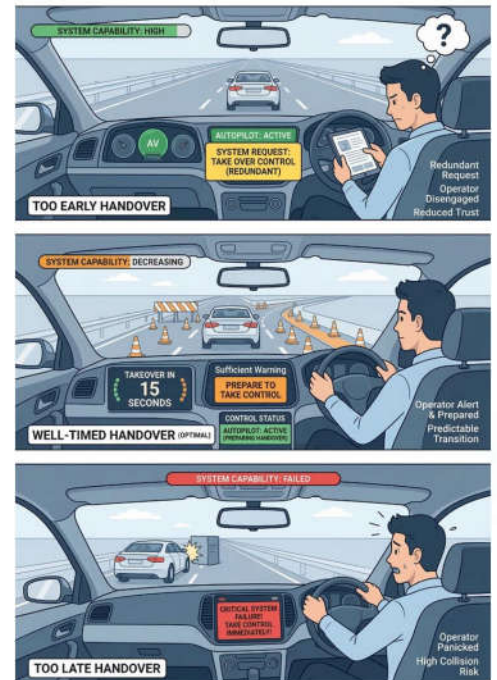
Abrupt vs Gradual Transitions

- Depends on risk level and urgency
 - **Abrupt Transitions**
 - **Immediate** change
 - Suitable for **emergencies**
 - **High urgency**
 - **Gradual Transitions**
 - **Progressive** visual cues
 - **Time** for adjustment
 - Suitable for **normal state changes**



Handover Timing

- Timing must balance **system capability** and **human readiness**
- Human takeover requires:
 - Sufficient warning time
 - Clear responsibility shift
 - Cognitive preparation
- If too **late**, reaction time is insufficient
- If too **early**, trust may decrease



Alerts

- Alerts are **not decorative** signals
- Mechanism to:
 - Capture **attention**
 - Signal **abnormal** conditions
 - Trigger **timely action**
- Poor alert can cause delay, confusion, or overreaction



Notification vs Warning vs Alarm

• Notification

- Informational
- No immediate action required



• Warning

- Potential risk
- Attention required
- Possible action needed



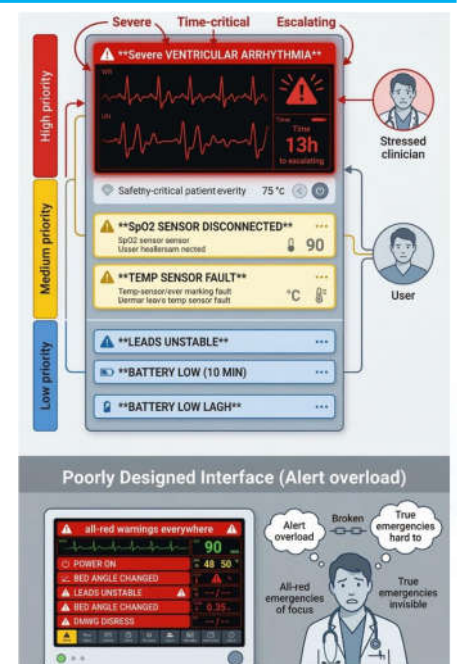
• Alarm

- Immediate danger
- Action required now



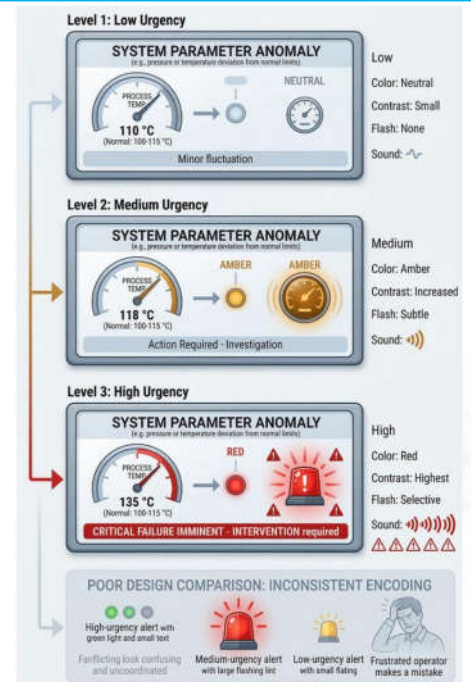
Prioritization

- If everything is urgent, nothing is urgent
 - Not all alerts are equally critical
- Prioritization should be based on:
 - Consequence Severity
 - Time Sensitivity
 - Escalation Probability



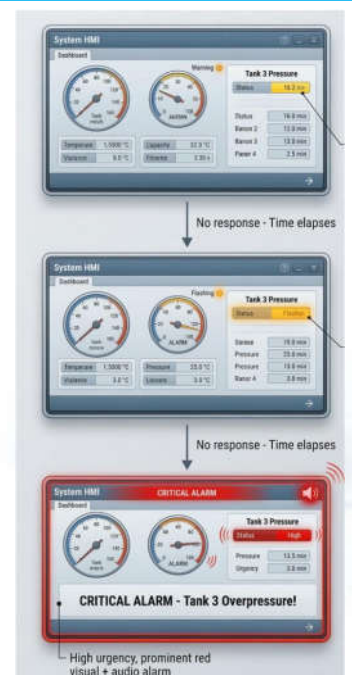
Urgency Encoding

- Urgency can be **encoded** through:
 - **Color** (e.g., neutral → amber → red)
 - **Size and contrast**
 - **Flashing or motion**
 - **Sound intensity and repetition**
- Must be **consistent** and easy to **interpret**



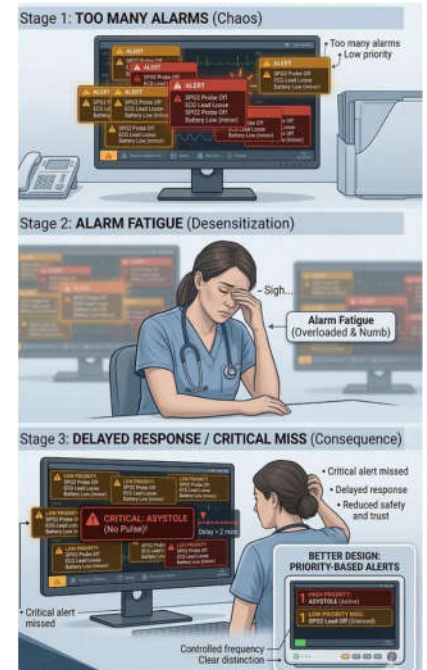
Escalation Strategies

- **Increasing visibility over time** if **no action** is taken
 - Supports timely intervention **without immediate overload**
- Examples:
 - **Static indicator** → **flashing signal**
 - **Visual only** → **visual + audio**
 - **Warning** → **alarm**



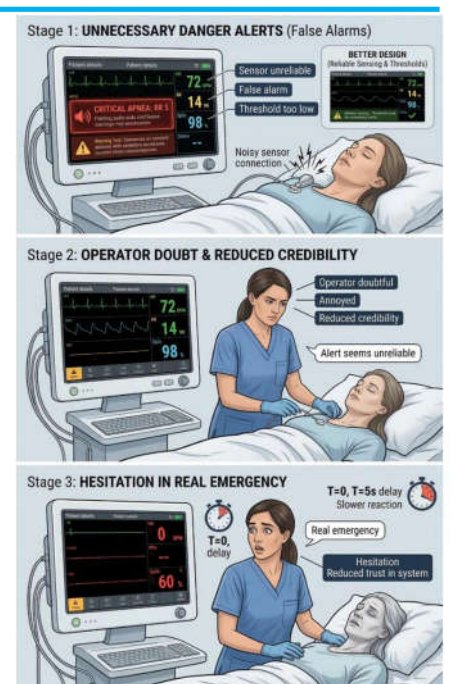
Alarm Fatigue

- **Alert frequency** must be carefully **controlled**
- Alarm **fatigue** occurs when:
 - **Too many** alarms are triggered
 - **Low-priority** alarms **dominate**
- **Consequences:**
 - **Ignored** critical alerts
 - **Delayed** response
 - **Loss of trust**



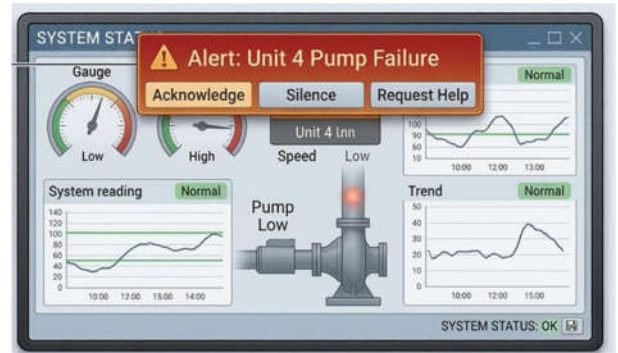
False Alarms

- Happen when:
 - System signals **danger unnecessarily**
 - **Thresholds** are **poorly set**
 - Sensors are **unreliable**
- **Effects:**
 - **Reduced credibility**
 - **Hesitation** during real emergencies



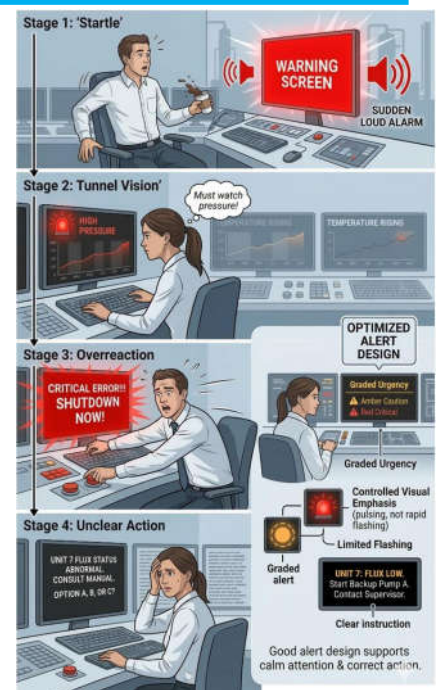
Acknowledgment Design

- Users must be able to:
 - Confirm the alert
 - Silence non-critical alarms
 - Escalate or request assistance
- Acknowledgment should:
 - Not hide unresolved problems
 - Clearly show current status



Human Factors Patterns

- Alert design must align with human cognitive limits
- Common human factors issues:
 - Startle effect from sudden alarms
 - Attention tunneling on flashing signals
 - Overreaction to loud warnings
 - Delay due to unclear instructions




Failures

• Hidden Modes

- Mode changes but the user does not notice
 - E.g. Aircraft autopilot mode changes without clear indication

• Delayed Feedback

- Slow responds without feedback
 - Users may
 - Repeat commands
 - Assume failure
 - Take unnecessary actions
 - Create conflicting inputs
- 


53

Failures

• Over-Automation Masking

- Automation can hide system complexity
 - Users may not understand what the system is doing
- Leads to automation complacency

• Poor State Visibility

- If system states are not clearly visible, users cannot maintain situational awareness
- 

54

UI Design

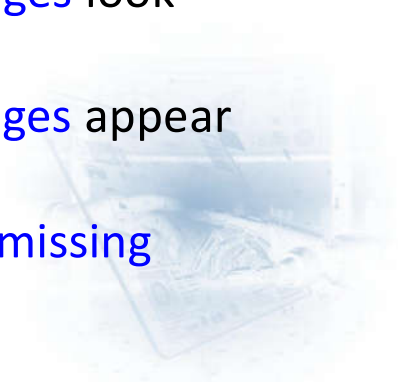
Failures

• Poor Prioritization

- Low-risk issues are displayed like high-risk
 - Lose urgency perception
 - Alarm fatigue
- Prioritization should reflect severity and time sensitivity

• Misleading Scales

- Axis scales are inconsistent
- Small changes look dramatic
- Large changes appear minor
- Context is missing



55

UI Design

Failures

• Missed Critical Alerts

- Critical alerts must dominate the visual field
- Design must ensure:
 - High visibility
 - Strong contrast
 - Clear differentiation

• Overlapping Alarms

- Multiple alarms trigger simultaneously
- Signals compete for attention
- Urgency levels are unclear



56

2D Screen Limitation

- **Traditional Interaction**

Environment ↔ **Display** ↔ Operator

- Screens **separated from the physical task environment**
- Problems:
 - Users must **shift gaze between task and display**
 - Information must be **mentally mapped** to real-world locations
 - **Spatial relationships** may be unclear



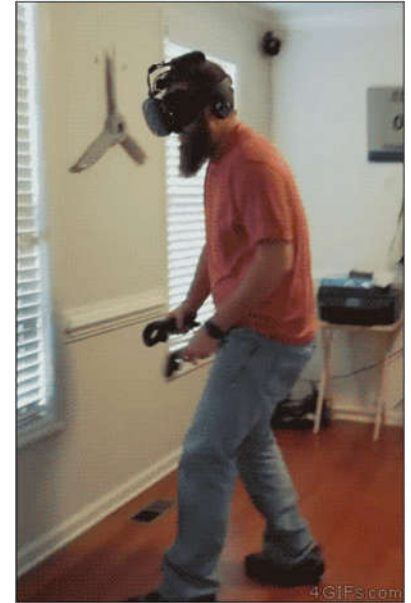
Immersive Interface

- Aim to **integrate digital content** with the **user's sensory perception** (vision, hearing, or touch), allowing users to interact with **virtual** or **augmented environments** in a more natural and spatial manner
 - Place **digital information** directly into the user's perception of the **environment**
- Extend traditional interfaces with different immersion and risk



Virtual Reality (VR)

- Creates a **fully simulated digital environment**
 - immersed through a **head-mounted display (HMD)**
 - **interact** using **controllers, gestures, or body motion**
 - Users **cannot see the real world** during the experience
- E.g. Surgical training, flight simulators



59

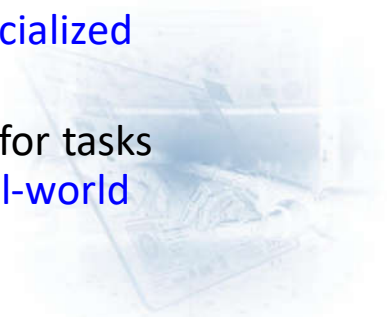
Virtual Reality (VR)

• Benefits

- **Fully immersive** interaction environment
- Supports **spatial understanding** of complex systems
- **Effective for simulation and education**
 - Enables safe training for dangerous tasks

• Limitations

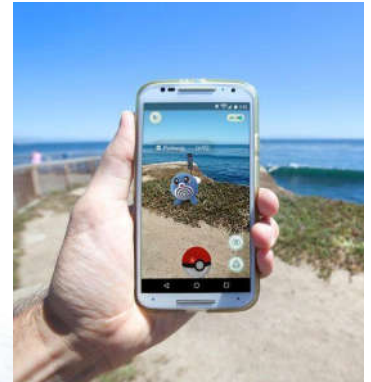
- Users **lose awareness** of the **real environment**
- **Motion sickness** and **visual fatigue** may occur
- Requires **specialized hardware**
- **Not suitable** for tasks requiring **real-world interaction**



60

Augmented Reality (AR)

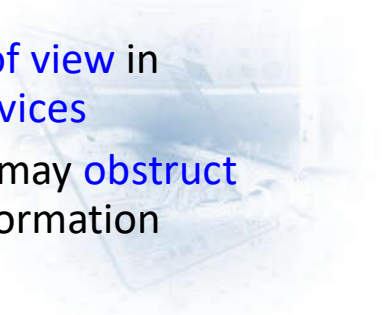
- **Overlays digital information on top of the real world,**
 - Usually through AR glasses or mobile devices
 - if head-mounted displays is used, it will be Mixed Reality
 - Users can see **both** the physical environment and digital elements simultaneously
- E.g. AR repair instructions, AR navigation systems



61

Augmented Reality (AR)

- | | |
|---|---|
| <ul style="list-style-type: none">• Benefits<ul style="list-style-type: none">• Combines real-world perception with digital guidance• Reduces cognitive translation between instructions and action• Improves situational awareness• Useful for maintenance, navigation, and training | <ul style="list-style-type: none">• Limitations<ul style="list-style-type: none">• Requires accurate environment tracking• Alignment errors between digital objects and real objects• Limited field of view in current AR devices• Visual clutter may obstruct real-world information |
|---|---|



62

Mixed Reality (MR)

- Blends **digital and physical worlds**, allowing virtual objects to **interact** with **real objects** and **respond to the environment**
 - Digital elements appear **anchored in the real environment** and can **change dynamically**
- E.g. Holographic models in engineering design.



63

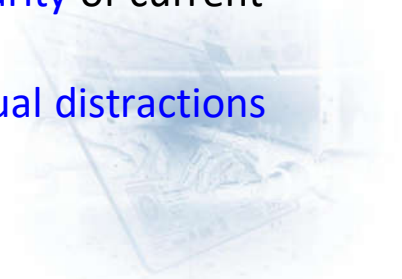
Mixed Reality (MR)

• Benefits

- All benefits of AR and VR
- Enables **natural interaction with digital objects**
- Supports **collaboration** in shared environments
- Useful for **design, engineering, and remote assistance**

• Limitations

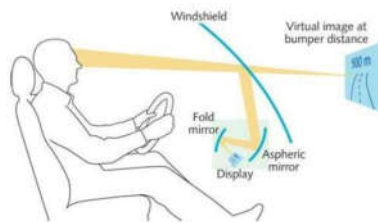
- **High computational and hardware requirements**
- **Tracking and environment mapping challenges**
- **Limited maturity of current devices**
- **Potential visual distractions**



64

Head-Up Display (HUD)

- Projects **essential information** into the **user's line of sight**,
 - allowing users to access information **without looking away** from their primary task
- E.g. Commonly used in vehicles and aircraft



Head-Up Display (HUD)

- **Benefits**
 - Reduces gaze shifts between display and environment
 - Supports **faster perception** of critical information
 - No wearable devices
 - Improves situational awareness
 - Useful for **navigation** and **safety alerts**
- **Limitations**
 - Limited display area
 - Too much information may cause **distraction**
 - **Brightness** and **visibility challenges** under sunlight
 - Risk of **visual clutter**

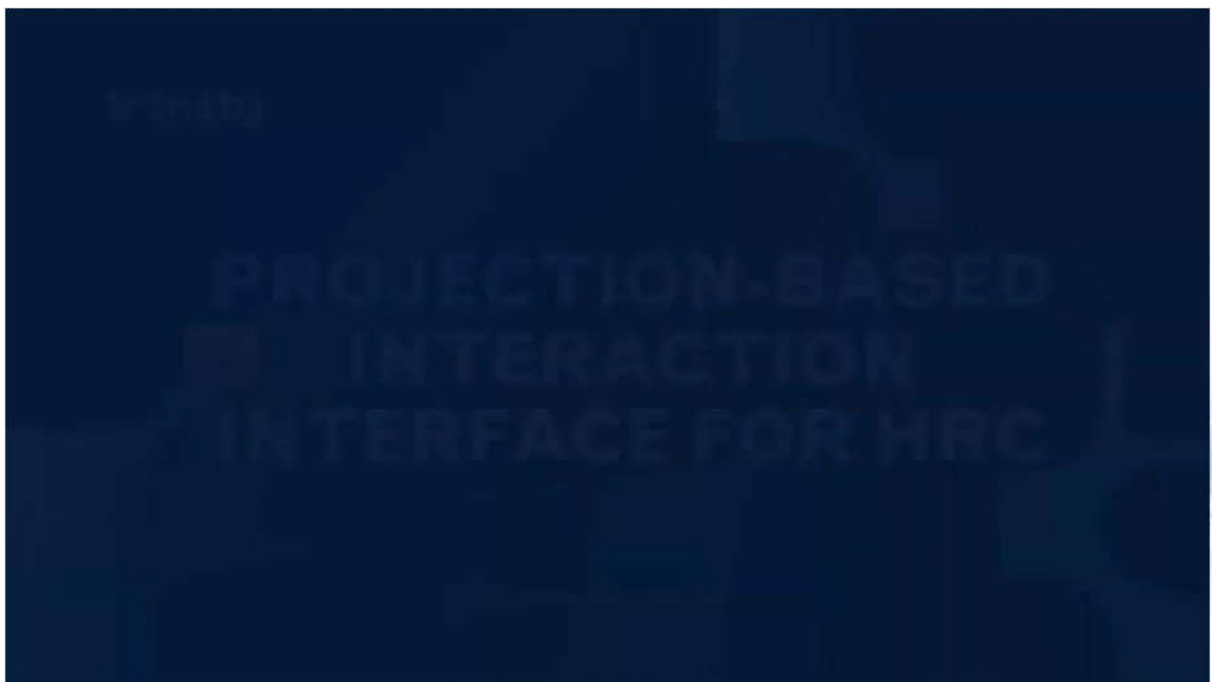


Projection-Based Interfaces

- Project digital information **directly onto real-world surfaces**
 - Users interact with the projected information through **gestures** or **touch**
- Examples: Interactive factory workstations, projection-based repair guidance



Projection-Based Interfaces



Projection-Based Interfaces

• Benefits

- Integrates digital information into the physical workspace
- Enables collaborative interaction
- Large display area
- No wearable devices
- Useful in industrial and educational environments

• Limitations

- Sensitive to lighting conditions
- Limited interaction precision
- Requires stable surfaces for projection
- Occlusion problems when users block the projection



Interface Comparison

Interface	Immersion Level	Ability	Risk
Virtual Reality	Very High	Fully virtual environment for training and simulation	Users lose awareness of the real world; motion sickness
Augmented Reality	Medium	Overlays digital information onto the real world	Misalignment and visual clutter
Mixed Reality	Medium–High	Digital objects interact with the real environment	Complex tracking; distraction
Head-Up Display	Low	Shows critical information in the user’s line of sight	Limited display space; information overload
Projection Interface	Low–Medium	Projects digital information onto physical surfaces	Lighting sensitivity; occlusion

Immersion Levels

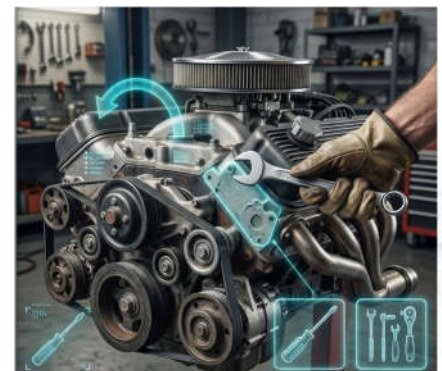
- **Higher Immersion \neq Better Performance**
 - Maintaining **awareness of the real world** is **essential** in **safety-critical systems**
- Excessive immersion can:
 - **Reduce awareness** of surrounding hazards
 - **Distract** users from the physical task
 - **Delay** response to unexpected events
- Choose the **appropriate level of immersion** based on the task and safety requirements



71

Benefits

- Immersive interfaces **align digital information with physical objects**
 - **Reduces** the need for **mental translation** between **information** and **action**
- E.g. **AR maintenance system**
 - Repair instructions appear directly on the machine component



72

Immersive Interface

Benefits

- **Reduce switch of attention** between the display and the system
 - Reduce the **workload** and **error**
- E.g. **Car Head-Up Display (HUD)**
 - Speed and navigation are shown on the windshield
 - The driver does not need to look down at the dashboard



73

Immersive Interface

Benefits

- **Reduce cognitive translation**
 - Faster understanding and response
 - Reduced cognitive load
- E.g. **Navigation system**
 - Driver sees a map and need to convert it to real-world direction
 - AR HUD: Arrow appears directly on the road ahead.



Immersive Interface

Benefits

- **Integrating system information with environmental perception**
- Examples
 - Highlighting hazards in AR
 - Visualizing robot movement paths
 - Displaying vehicle sensor information



Immersive Interface

Interaction Method

- Each method has **trade-offs in precision, fatigue, and reliability**:
 - **Gaze-based** selection
 - **Gesture** control
 - **Voice** commands
 - **Hand** controllers
 - Physical **buttons**



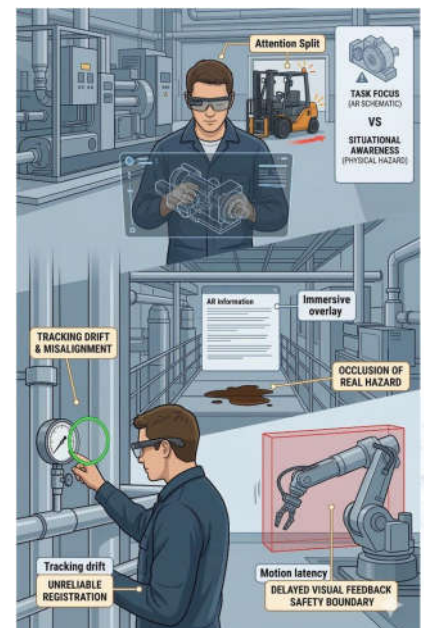
Hardware Constraints

- Immersive systems face **physical limitations**:
 - Field of view size
 - Display brightness in outdoor environments
 - Tracking accuracy
 - Latency
 - Weight and ergonomics
- **Hardware constraints** directly affect usability and safety



Human Factors

- Unlike traditional screens, immersive interfaces:
 - Share visual space with the real world
 - Compete with physical hazards
 - Depend on tracking and motion
- Will new human factors risks be introduced?



Occlusion

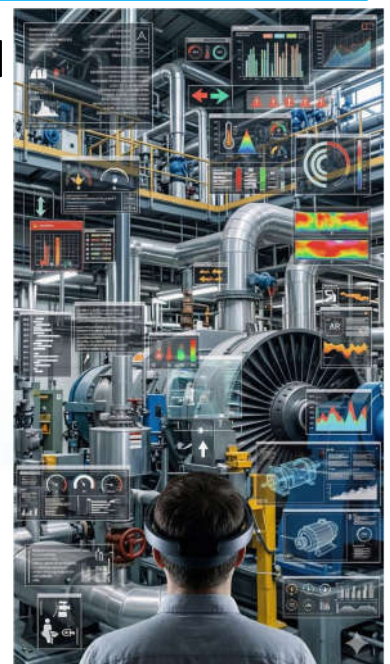
- Digital elements **block** real-world objects
 - Hide critical environmental cues
 - Reduce peripheral awareness
- In safety-critical contexts
 - Delay hazard detection
 - Create false confidence



79

Visual Clutter

- **Minimalism** is **essential** in immersive HMI
- **Too many overlays** are displayed
 - Multiple layers compete for attention
 - Contrast and hierarchy are weak
- Effects:
 - Slower decision-making
 - Increased cognitive load
 - Missed critical information



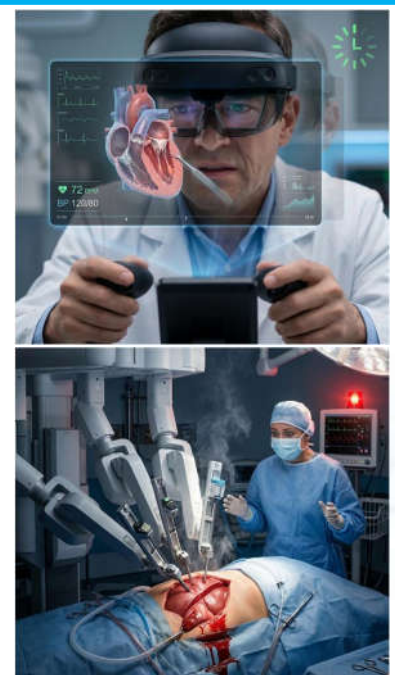
Registration Errors

- Digital overlays **do not align with physical objects**
 - Tracking drifts over time
 - Motion causes instability
- Consequences:
 - Misinterpretation of position
 - Dangerous action based on incorrect alignment



Delayed response

- **Latency** may occur due to **sensor, network, graphics rendering, system processing**
- Delayed visual feedback can cause:
 - incorrect perception of object position
 - mismatch between real movement and displayed information
 - slower user reaction and decision-making



Motion Sickness & Fatigue

- **Low latency** and **ergonomic design** are **mandatory**
- Immersive systems **may cause discomfort** due to:
 - **Latency** between **head movement** and **display update**
 - **Mismatch** between **visual motion** and **body sensation**



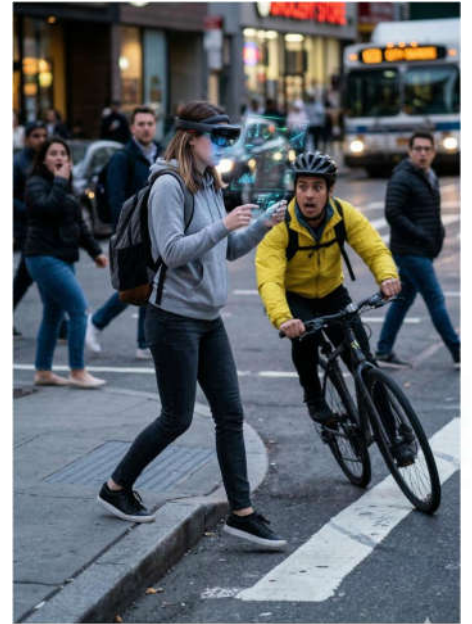
Overtrust in Overlays

- Highlighted digital **cues** may **appear authoritative**
- Users may:
 - **Follow** overlays **without verification**
 - **Ignore conflicting** real-world signals
 - **Over-rely** on automation
- Immersive design must **avoid creating blind trust**



Attention Diversion

- Overlays can:
 - Pull attention away from hazards
 - Narrow focus excessively
 - Reduce peripheral awareness
- Attention must remain **balanced between: Digital information and physical environment**



Context Switching

- Operators may **switch** between:
 - Immersive **view**
 - Physical **controls**
 - External **displays**
- **Frequent switching** increases:
 - Cognitive **load**
 - **Error** probability
- Immersive interfaces **should reduce, not increase, context switching**



Design Principles

- Engineering rules
 - Goal: **maximize** safety and reliability
 - Goal: **not maximum** immersion
- Design must prioritize:
 - Safety → Clarity → Stability → Trust



Real-World Dominance

- **Physical environment** must remain **primary**
- **Digital information** should **support reality**
- Guidelines:
 - Do not block critical objects
 - Preserve peripheral vision
 - Avoid full-screen overlays in operational tasks



Minimal Overlays

- Display **only**:
 - Task-relevant information
 - Time-critical signals
 - Actionable guidance
- **Avoid**:
 - Decorative graphics
 - Excessive annotations
 - Persistent non-essential elements



Priority Hierarchy

- Use **hierarchy** to indicate:
 - Critical vs secondary
 - Immediate vs future
 - Risk vs normal state
- Hierarchy must be instantly recognizable
- Priority should be encoded (size, motion, position)



Adaptive Visibility

- **Adapt display** based on:
 - Task phase
 - Environmental condition
 - User workload
 - Risk level
- **Low-risk state** → simplified display
- **High-risk state** → increased guidance



91

Failure Fallback

- Tracking or sensors **may fail**
- **Never leave incorrect information visible**
 - Detect failure clearly
 - Remove misleading overlays
 - Notify the operator immediately
- Fallback behavior must be **explicit**



92

Safe Degradation

- When system **capability reduces**:
 - Reduce automation level
 - Simplify interface
 - Maintain core functionality
- Graceful degradation **prevents sudden confusion**
- **Avoid abrupt loss** of essential information



Trust Calibration

- Immersive interfaces **influence trust**
- Design should:
 - Show uncertainty
 - Avoid false precision
 - Make automation boundaries **visible**
 - Indicate **confidence** levels



Audio

- Vision is dominant, but **audio** is **critical**
- Audio is **not** secondary decoration
It is a **functional safety channel**
- The auditory channel:
 - Complements visual information
 - Captures attention quickly
 - Supports operation in **dynamic** environments



95

Audio

Complementary Modality

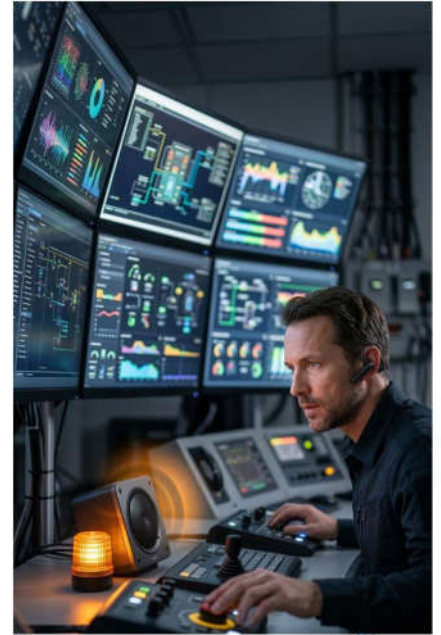
- Audio **complements vision** by:
 - Providing **redundancy** for critical **alerts**
 - **Reducing visual overload**
 - When **visual** attention is **limited**, audio maintains awareness
 - Supporting **multimodal** feedback
 - Proper coordination between vision and audio improves robustness



96

Eyes-Busy Scenarios

- Audio is **especially valuable** when:
 - The operator must **focus visually** on the environment
 - **Hands** are **occupied**
 - **Movement** is **required**
- Audio supports safe operation
 - E.g. Driving, Surgery, Industrial maintenance



Temporal Advantages

- Sound has **strong temporal properties**:
 - **Immediate attention** capture
 - **Fast** perception
 - **Effective** for **urgency** signaling
 - Does **not** require **visual** fixation
- Unlike visuals, **audio can interrupt ongoing tasks** effectively



Urgency Encoding

- Which sound feels more urgent?



Sound A



Sound B



Sound C



Sound D



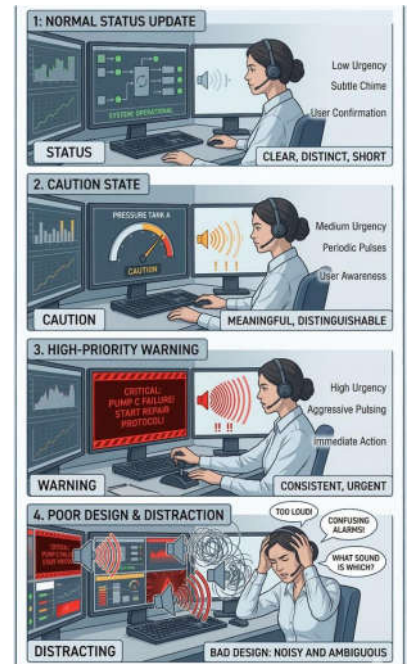
Urgency Encoding

- Urgency can be conveyed through:
 - **Pitch** (higher = higher urgency)
 - **Tempo** (faster = higher urgency)
 - **Rhythm** (faster = higher urgency)
 - **Volume** (larger = higher urgency, used carefully!)
- Encoding must be **consistent** and easy to **distinguish**



Effective Feedback

- Sound should **communicate state and urgency without causing distraction**
- Sound Feedback must be:
 - Clear
 - Distinguishable
 - Meaningful
 - Consistent



Meaning of Sound

- Users **should not need to guess** what a sound means

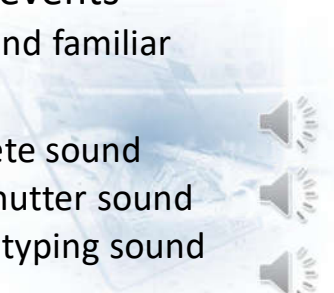
• Method:

• **Earcons**

- Meaning **learned** through training
- Abstract tones or musical patterns
- Example:
 - Windows system notification sound

• **Auditory Icons**

- **Real-world sounds mapped to events**
 - Intuitive and familiar
- Example
 - Trash delete sound
 - Camera shutter sound
 - Keyboard typing sound





Meaning of Sound

- **Effective mapping** requires:
 - Clear association between sound and event
 - Consistency across system states
 - Distinct sounds for distinct meanings
- **Avoid:**
 - Similar tones for different alerts
 - Overly complex sound patterns



103

Escalation via Sound

- Escalation **supports gradual attention capture** without immediate overload
- Escalation strategy:
 - Low-priority notification → short tone 
 - Warning → repeated tone 
 - Alarm → louder, faster repetition 
- Sound intensity **should match risk level**



104

Directionality

- Humans can **localize sound sources**
- Directional sound can **guide attention without visual scanning**
- Audio can provide:
 - **Spatial orientation**
 - **Directional warnings**
 - **Source identification**



105

When Audio Outperforms Visual

- In time-critical situations, sound can **reduce response delay**
- Audio is superior when:
 - **Immediate reaction** is required
 - **Visual** channel is **overloaded**
 - **Visibility** conditions are **poor**
 - **Continuous monitoring** is impractical



106

Overload

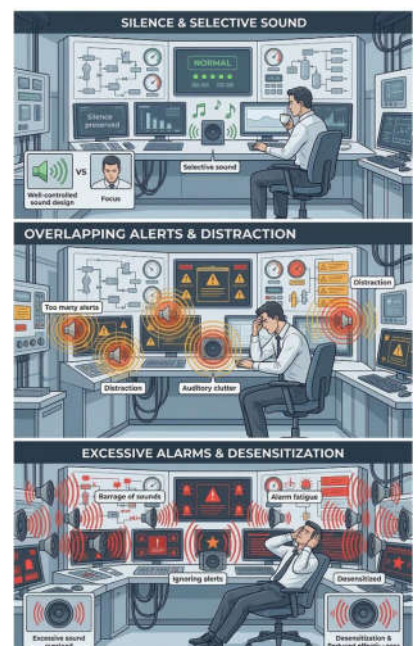
- Too many sounds cause **stress** and **desensitization**
 - Sound alerts may be ignored
- Design Implication:
 - Limit number of **distinct sounds**
 - Reserve audio for **important** events
 - Avoid **continuous** or **unnecessary** tones



Audio

Risks of Overuse

- **Silence** can be **as important** as sound
- Too many sounds reduce effectiveness.
 - Annoyance
 - Distraction
 - Alarm fatigue
 - Desensitization



Redundancy

- **Critical events** should use be **alerted** by **both visual** and **auditory** signals
 - **Redundant signals** must be **synchronized** and **non-conflicting**
- Redundancy improves reliability



109

Speech Interaction

- One of the **most natural forms of human communication**
- Advantages
 - **Natural** communication (no and little training)
 - **Hands-free** interaction
 - **Eyes-free** interaction
 - **Faster input** for some tasks



110

Components

- Speech interaction typically includes several stages:

- 1. Speech Recognition**

Spoken Audio → Text

- 2. Natural Language Understanding (NLU)**

Interprets meaning

- 3. Dialogue Management**

Determines how to response

- 4. Speech Synthesis (Text-to-Speech)**

Text → Spoken Output



111

Types

- **Command-Based Interaction**

- Speak **predefined commands**.
 - E.g. “Turn on the lights.”
- **Advantages**
 - **Reliable**
 - **Easy to interpret**
- **Limitations**
 - **Limited flexibility**

- **Conversational Interaction**

- Communicate using **natural language**
 - E.g. “What should I do?”
- **Advantages**
 - More **natural** interaction
 - **Supports complex queries**
- **Limitations**
 - **Higher** system complexity

112

User Frustration

- Speech interaction can frustrate users when:
 - Commands must be repeated
 - Misunderstands frequently
 - Responses are unnatural or rigid
- Frustration reduces compliance and increases manual override
 - Should focus on Turn-Taking, Determination, Interruption Handling, Feedback Timing, Confirmation Strategy, Error Recovery



113

Turn-Taking

- Requires clear **control of who is speaking**
- Challenges
 - Interruptions
 - Overlapping speech
 - Delayed system responses
- Design should provide signals such as:
 - Listening indicators
 - Audio feedback
 - Confirmation prompts



114

Interruptions

- **Interruptions** are **inevitable**
 - User corrects system
 - System and User talk at the same time
 - Multiple commands overlap
 - Emergency situations override normal dialogue
- **Interruption management** is important to avoid:
 - Confusion
 - Missed instructions
 - Unsafe outcomes



115

Feedback Timing

- **Timing** is part of interaction **quality**
 - Immediate acknowledgment
 - Processing indication
 - Clear response
- **Delayed feedback** may cause:
 - Repeated commands
 - User confusion
 - Loss of trust



116

Confirmation Strategy

- **Critical commands** require **confirmation**
 - E.g. “Confirm shutdown?”
 - Prevents unintended actions
- However, **excessive confirmation slows** workflow
- **Tradeoff** between safety and efficiency



117

Error Recovery

- **Speech errors** are **inevitable**
- **Error handling** must be smooth and predictable.
 - Repetition prompts
 - Clarification questions
 - Alternative suggestions
 - Safe cancellation



118

Over-Expectation

- Human-like voice can **create unrealistic expectations**
 - Users may **assume more than the system can do**:
 - **Deep understanding**
 - **Full reasoning ability**
 - **Emotional awareness**
- **Mismatch** between expectation and capability leads to **frustration** and **misuse**.
- **Clearly** communicate system **limitations**



119

Limitation

- Speech interactions are **less visible than graphical interfaces**
 - **Cannot** easily **see available options**
 - Users have to **remember** the commands



120

Challenges

- **Language Interpretation**
 - Ambiguity
 - Accent and pronunciation variation
- **Environment Issues**
 - Noisy environments
 - Multiple speakers
 - Echo and reverberation
- **Advanced AI**, e.g. LLM, may **help** but **high cost**



121

Safety Constraints

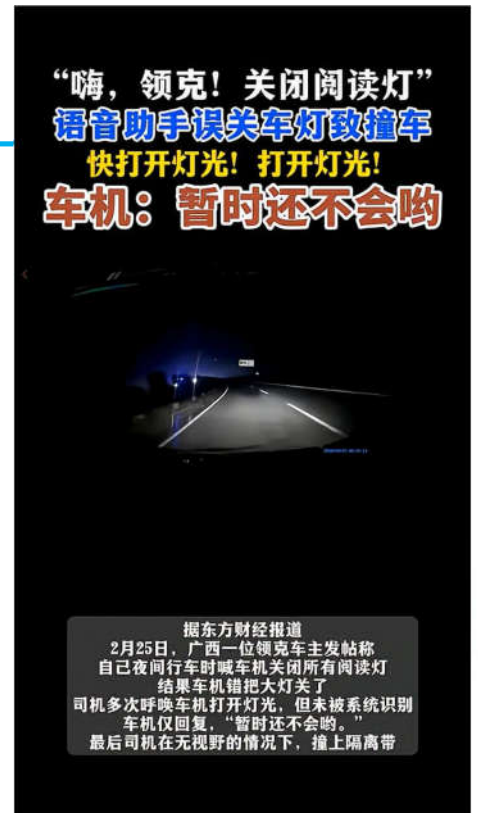
- In safety-critical systems:
 - Speech should **not** be the **sole control method**
 - **Critical commands** require **confirmation**
 - **Fail-safe alternatives** must exist
- Voice control must **never reduce system safety**
- **Redundancy** and clear **boundaries** are **essential**



122

Safety Constraints

- E.g. **Voice Misrecognition** Caused a **Highway Crash** in Guangxi Expressway
 - Lynk & Co Z20, February 25, 2026
 - The driver wanted to use a voice command to turn off the interior reading lights
 - However, the machine turned off all vehicle lights, including the headlights
 - The driver tried to use voice control again to turn the lights back on, but the system failed to recognize the request correctly and reportedly replied, "I still don't know how."



Multimodality

Human Sensory Integration

- Humans **naturally integrate multiple senses**
- Multimodal interaction **aligns with natural perception processes**
- **Well-designed** multimodal systems **feel intuitive and robust**



Why Not Single Modality?

- **Single-modality is fragile**
 - Bright sunlight reduces screen visibility
 - Loud environments reduce speech clarity
 - Physical movement limits visual attention
- **Multimodal systems adapt to environmental constraints**



Why Not Single Modality?



Multimodal Fusion

- **Combining** information from multiple channels to support perception, decision, and action
- There are two types:
 - **Complementary**
 - Each modality provides **different** information
 - Increases richness
 - Example: **visual** path + **auditory** warning
 - **Redundant**
 - **Same** critical information delivered via multiple channels
 - Increases safety
 - Example: **flashing red** + **alarm** sound

Complementarity

- **Combining** multiple modalities allows one channel to **compensate** for the **limitations** of another
- Example
 - **Vision**
 - **Pro:** **Spatial** information
Persistent display
High data bandwidth
 - **Con:** **Overload**
Occlusion
Limited field of view
 - **Audio**
 - **Pro:** **Temporal** alerts
Attention capture
Directional cues
 - **Con:** **Noise**
Desensitization
Limited capacity



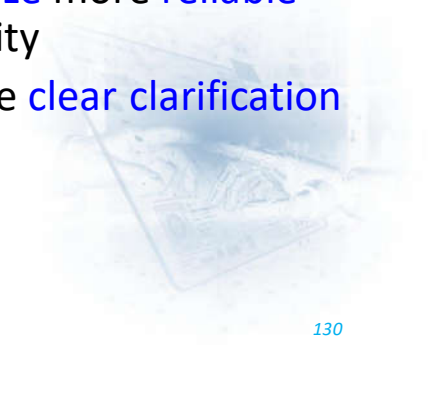
Redundancy

- **Critical information** can be presented through **different channels**
 - If **one** channel **fails**, **another** remains **active**
- Redundancy **increases**:
 - **Detection** probability
 - **Reliability**
 - Reaction **speed**



Design Principle

- **Temporal Synchronization**
 - **Timing alignment** is critical
 - E.g. **Confusion** when
 - **Visual** alert appears **before** sound
 - **Sound** triggers **without** **visual** confirmation
- **Conflict Resolution**
 - **Multimodality** disagrees
 - System must:
 - **Detect** inconsistencies
 - **Prioritize** more **reliable** modality
 - Provide **clear clarification**



Design Principle

• Modality Dominance

- Dominance should match urgency
- Maintain balanced awareness in normal situations

• Reliability-Based Weighting

- Not all modalities are equally reliable
- Emphasize more reliable channel
- Adaptive reliability improves robustness

131

Design Principle

• Attention Management

- Multi-signal competes for limited attention
- Should:
 - Avoid simultaneous overload
 - Sequence signals logically
 - Escalate gradually
 - Prevent startle effects

• Adaptive Modality Switching

- Adapt modality based on:
 - Task phase
 - User workload
 - Environmental condition
 - Risk level
- Example:
 - Low risk: Visual only
 - High risk: Visual + audio

132

Failure Handling

- **Failure-aware fusion** increases safety
 - When one modality **fails**:
 - **Detect** the failure
 - **Notify** the user
 - **Increase reliance** on remaining modalities
- **Never** leave users **unaware** of **degraded capability**

