

**Toward Ergonomic and Productive Human-Machine Interaction: A  
Conceptual Study of Hybrid Gesture and Voice Interfaces.**

Submitted By

**Mapon Mounde Mariama Sheila**

200011154

**Abdoul Wadoud Salahouddine**

200011161

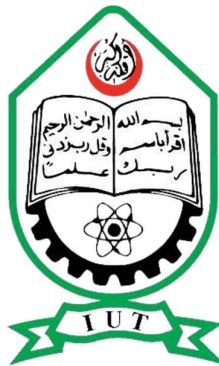
**Moufadhul Said Bacar**

200011143

Supervised By

**Dr. Md Abu Shaid Sujon**

**A Thesis submitted in partial fulfillment of the requirement for the degree  
of Bachelor of Science in Mechanical Engineering**



**Department of Mechanical and Production Engineering (MPE)**

**Islamic University of Technology (IUT)**

**October, 2025**

### **Candidate's Declaration**

This is to certify that the work presented in this thesis, titled, "Toward Ergonomic and Productive Human-Machine Interaction: A Conceptual Study of Hybrid Gesture and Voice Interfaces", is the outcome of the investigation and research carried out by me under the supervision of Dr. Md Abu Shaid Sujon ,Assistant Professor.

It is also declared that neither this thesis nor any part of it has been submitted elsewhere for the award of any degree or diploma.

-----  
Mapon Mounde Mariama Sheila  
Student No: 200011154

-----  
Abdoul Wadoud Salahouddine  
Student No: 200011161

-----  
Moufadhul Said Bacar  
Student No:200011143

## **Recommendation of the Thesis Supervisors**

---

The thesis titled “Toward Ergonomic and Productive Human–Machine Interaction: A Conceptual Study of Hybrid Gesture and Voice Interfaces” submitted by Mapon Mounde Mariama Sheila, Abdoul Wadoud Salahouddine, Moufadhul Said Bacar Student No: 200011154,200011161,200011154 respectively ,has been accepted as satisfactory in partial fulfillment of the requirements for the degree of BSc. in Mechanical Engineering **on 06<sup>th</sup>October,2025.**

1. -----

Dr. Md Abu Shaid Sujon(Supervisor)

Assistant Professor

Mechanical and Production Engineering(MPE) Department,IUT, Board

Bazar, Gazipur-1704, Bangladesh.

## CO-PO Mapping of ME 4800 -Thesis and Project

| COs  | Course Outcomes (CO) Statement  | (PO)                                   | Addressed by                |   |
|------|---|--|-----------------------------|---|
| CO 1 | Discover and Locate research problems and illustrate them via figures/tables or projections/ideas through field visit and literature review and <u>determine/Setting</u> aim and objectives of the project/work/research in specific, measurable, achievable, realistic and timeframe manner. | PO2<br>Problem analysis                | Thesis Book                 | √ |
|      |   |  | Performance by research     | √ |
|      |   |  | Presentation and soft skill | √ |
| CO 2 | Design research solutions of the problems towards achieving the objectives and its application. Design systems, components or processes that meets related needs in the field of mechanical engineering   | PO3<br>Design/development of solutions | Thesis Book                 | √ |
|      |   |  | Performance by research     | √ |
|      |   |  | Presentation and soft skill | √ |
| CO 3 | Review, debate, compare and contrast the relevant literature contents. Relevance of this research/study. Methods, tools, and techniques used by past researchers and justification of use of them in this work.   | PO4<br>Investigation                   | Thesis Book                 | √ |
|      |   |  | Performance by research     | √ |
|      |   |  | Presentation and soft skill | √ |
| CO 4 | Analyze data and exhibit results using tables, diagrams, graphs with their interpretation. Investigate the designed solutions to solve the problems through case study/survey study/experimentation/simulation using modern tools and techniques.   | PO5<br>Modern tool usage               | Thesis Book                 | √ |
|      |   |  | Performance by research     | √ |
|      |   |  | Presentation and soft skill | √ |
| CO 5 | Apply moral values and research/professional ethics throughout the work, and justify genuine referencing on sources, and demonstration of own contribution.   | PO8<br>Ethics                          | Thesis Book                 | √ |
|      |   |  | Performance by research     | √ |
|      |   |  | Presentation and soft skill | √ |
| CO 6 | Perform own self and manage group activities from the beginning to the end of the research/work as a quality work.  | PO9<br>Individual work and teamwork    | Thesis Book                 | √ |
|      |   |  | Performance by research     | √ |
|      |   |  | Presentation and soft skill | √ |
| CO 7 | Compile and arrange the work outputs, write the report/thesis, a sample journal paper, and present the work to wider audience using modern communication tools and techniques.  | PO10<br>Communication                  | Thesis Book                 | √ |
|      |   |  | Performance by research     | √ |
|      |   |  | Presentation and soft skill | √ |
| CO 8 | Recognize the necessity of life-long learning in career development in dynamic real-world situations from the experience of completing this project.  | PO12<br>Life-long learning             | Thesis Book                 | √ |
|      |   |  | Performance by research     | √ |
|      |   |  | Presentation and soft skill | √ |

Student Name /ID:

1. Mapon Mounde Mariama Sheila(200011154)
2. Abdoul Wadoud Salahouddine(200011161)
3. Moufadhul Said Bacar(200011143)

Signature of the Supervisor:

.....

Name of the Supervisor:  
Dr. Md. Abu Shaid Sujon

**K-P-A Mapping of ME 4800 -Theis and Project**

| COs | POs  | Related Ks |    |    |    |    |    |    |    | Related Ps |    |    |    |    |    |    | Related As |    |    |    |    |
|-----|------|------------|----|----|----|----|----|----|----|------------|----|----|----|----|----|----|------------|----|----|----|----|
|     |      | K1         | K2 | K3 | K4 | K5 | K6 | K7 | K8 | P1         | P2 | P3 | P4 | P5 | P6 | P7 | A1         | A2 | A3 | A4 | A5 |
| CO1 | PO2  | ✓          | ✓  | ✓  | ✓  |    |    |    |    | ✓          |    |    |    |    |    |    |            |    |    |    |    |
| CO2 | PO3  |            |    |    |    | ✓  |    |    |    | ✓          | ✓  | ✓  |    |    |    | ✓  | ✓          | ✓  | ✓  |    |    |
| CO3 | PO4  |            |    |    |    |    |    |    | ✓  | ✓          | ✓  | ✓  |    |    |    | ✓  | ✓          |    |    |    |    |
| CO4 | PO5  |            |    |    |    |    | ✓  |    |    | ✓          |    |    |    | ✓  |    |    |            |    |    |    |    |
| CO5 | PO8  |            |    |    |    |    |    | ✓  |    |            |    |    |    |    |    |    |            |    |    |    |    |
| CO6 | PO9  |            |    |    |    |    |    |    |    |            |    |    |    |    |    |    |            |    |    |    |    |
| CO7 | PO1  |            |    |    |    |    |    |    |    |            |    |    |    |    |    |    |            |    |    |    |    |
| CO8 | PO12 |            |    |    |    |    |    |    |    |            |    |    |    |    |    |    |            |    |    |    |    |

Student Name /ID:

Signature of the Supervisor:

1 Mapon Mounde Mariama Sheila

.....

2. Abdoul Wadoud Salahouddine

Name of the Supervisor: Dr.  
Md Abu Shaid Sujon

3. Moufadhul Said Bacar

---

## List of Sustainable Development Goals (SDGs) Addressed in this Project

| SDG No.  | Goals       | Targets   | Relevance to the Thesis    | Remarks                                     |
|--|-------------|---|----------------------------|---|
| 1  | No Poverty  | 1.1 Eradicate extreme poverty (people living on less than \$1.25/day).  |                            |   |
|  |             | 1.2 Reduce poverty in all its forms by at least half.   |                            |   |
|  |             | 1.3 Implement nationally appropriate social protection systems.   |                            |   |
|  |             | 1.4 Ensure equal rights to economic resources, services, property, inheritance, technology, and financial services. |                            |   |
|  |             | 1.5 Build resilience of the poor and reduce exposure to climate-related and other shocks.                           |                            |   |
|  |             | 1.a Mobilize resources to end poverty.  |                            |   |
|  |             | 1.b Create pro-poor policy frameworks.  |                            |   |
| 2  | Zero Hunger | 2.1 End hunger and ensure access to safe, nutritious food year-round.   |                            |   |
|  |             | 2.2 End all forms of malnutrition.  |                            |   |
|  |             | 2.3 Double agricultural productivity and incomes of small-scale producers.  |                            |   |
|  |             | 2.4 Ensure sustainable food production systems and resilient agricultural practices.                                |                            |   |
|  |             | 2.5 Maintain genetic diversity of seeds, plants, and animals.   |                            |   |
|  |             | 2.a Increase investment in rural infrastructure, research, and technology.  |                            |   |
|  |             | 2.b Correct and prevent trade restrictions/distortions in global food markets.                                      |                            |   |
|  |             | 2.c Adopt measures to ensure proper functioning of food commodity markets.  |                            |   |
|  |             | 3   | Good Health and Well-Being | 3.1 Reduce global maternal mortality ratio. |
| 3.2 End preventable deaths of newborns and under-5 children.                       |             |   |                            |   |
| 3.3 End epidemics of AIDS, tuberculosis, malaria, and neglected tropical diseases. |             |   |                            |   |

|   |                   |   |   |   |
|---|-------------------|---|---|---|
|   |                   | 3.4 Reduce premature mortality from NCDs and promote mental health.                 |   |   |
|   |                   | 3.5 Strengthen prevention and treatment of substance abuse.                         |   |   |
|   |                   | 3.6 Halve global deaths/injuries from road traffic accidents.                       |   |   |
|   |                   | 3.7 Ensure universal access to sexual and reproductive healthcare.                  |   |   |
|   |                   | 3.8 Achieve universal health coverage.  | √ | The hybrid HMI reduces physical strain and cognitive fatigue, contributing to a healthier and safer work environment. |
|   |                   | 3.9 Reduce deaths from hazardous chemicals, pollution, and contamination.           |   |   |
|   |                   | 3.a Strengthen tobacco control (WHO FCTC).  |   |   |
|   |                   | 3.b Support R&D of vaccines and medicines.  |   |   |
|   |                   | 3.c Increase health financing and workforce.  |   |   |
|   |                   | 3.d Strengthen capacity for early warning and risk management.                      |   |   |
| 4 | Quality Education | 4.1 Ensure all complete free, equitable, quality primary and secondary education.   |   |   |
|   |                   | 4.2 Ensure access to quality early childhood development and pre-primary education. |   |   |
|   |                   | 4.3 Ensure equal access to affordable technical, vocational, and higher education.  |   |   |
|   |                   | 4.4 Increase skills for employment and entrepreneurship.                            |   |   |
|   |                   | 4.5 Eliminate gender disparities in education.                                      |   |   |
|   |                   | 4.6 Ensure literacy and numeracy for youth and adults.                              |   |   |
|   |                   | 4.7 Ensure learners acquire knowledge/skills for sustainable development.           |   |   |
|   |                   | 4.a Build and upgrade education facilities that are inclusive and safe.             |   |   |
|   |                   | 4.b Expand scholarships for developing countries.                                   |   |   |

|   |                             |  |  |  |
|---|-----------------------------|--|--|--|
|   |                             | 4.c Increase supply of qualified teachers.                                   |  |  |
| 5 | Gender Equality             | 5.1 End all forms of discrimination against women and girls.                 |  |  |
|   |                             | 5.2 Eliminate violence against women and girls.                              |  |  |
|   |                             | 5.3 Eliminate harmful practices (child, early, forced marriage, FGM).        |  |  |
|   |                             | 5.4 Recognize and value unpaid care and domestic work.                       |  |  |
|   |                             | 5.5 Ensure women's participation in leadership and decision-making.          |  |  |
|   |                             | 5.6 Ensure universal access to reproductive health and rights.               |  |  |
|   |                             | 5.a Undertake reforms to give women equal rights to resources.               |  |  |
|   |                             | 5.b Enhance use of enabling technology to empower women.                     |  |  |
|   |                             | 5.c Adopt and strengthen policies and laws for gender equality.              |  |  |
| 6 | Clean Water and Sanitation  | 6.1 Achieve universal and equitable access to safe drinking water.           |  |  |
|   |                             | 6.2 Achieve access to adequate sanitation and hygiene.                       |  |  |
|   |                             | 6.3 Improve water quality by reducing pollution.                             |  |  |
|   |                             | 6.4 Increase water-use efficiency and sustainable withdrawals.               |  |  |
|   |                             | 6.5 Implement integrated water resources management.                         |  |  |
|   |                             | 6.6 Protect and restore water-related ecosystems.                            |  |  |
|   |                             | 6.a Expand international cooperation in water and sanitation.                |  |  |
|   |                             | 6.b Support participation of local communities.                              |  |  |
| 7 | Affordable and Clean Energy | 7.1 Ensure universal access to affordable, reliable, modern energy services. |  |  |
|   |                             | 7.2 Increase substantially the share of renewable energy.                    |  |  |

|   |                                 |  |   |   |
|---|---------------------------------|--|---|---|
|   |                                 | 7.3 Double global rate of improvement in energy efficiency.                              |   |   |
|   |                                 | 7.a Enhance international cooperation on clean energy research/technology.               |   |   |
|   |                                 | 7.b Expand infrastructure and upgrade technology for sustainable energy.                 |   |   |
| 8 | Decent Work and Economic Growth | 8.1 Sustain per capita economic growth.  | √ | By improving operator efficiency and productivity, the system supports sustainable economic growth.                   |
|   |                                 | 8.2 Achieve higher levels of productivity through diversification, tech, and innovation. | √ | The interface achieves higher productivity through technological innovation and diversification of interaction modes. |
|   |                                 | 8.3 Promote policies for decent job creation and entrepreneurship.                       | √ | It promotes policies for decent job creation by enhancing the quality and ergonomics of work.                         |
|   |                                 | 8.4 Improve resource efficiency in production and consumption.                           | √ | The conceptual design approach improves resource efficiency by optimizing the human-system interaction process.       |
|   |                                 | 8.5 Achieve full and productive employment for all.                                      |   |   |
|   |                                 | 8.6 Substantially reduce youth not in employment/education/training.                     |   |   |
|   |                                 | 8.7 Eradicate forced labour, modern slavery, and child labour.                           |   |   |
|   |                                 | 8.8 Protect labour rights and safe working environments.                                 | √ | The system protects labour rights by  |

|  |  |  |                      |   |
|--|--|--|----------------------|---|
|  |  |  |                      | fostering safe, ergonomic, and less frustrating working environments.                               |
|  |  | 8.9 Promote sustainable tourism.   |                      |   |
|  |  | 8.a Increase aid for trade support.  |                      |   |
|  |  | 8.b Develop a global youth employment strategy.                                      |                      |   |
| 9  | Industry, Innovation, and Infrastructure | 9.1 Develop quality, reliable, sustainable infrastructure.                           | √                    | It develops a conceptual framework for a sustainable and quality human-machine infrastructure.      |
|  |  | 9.2 Promote inclusive and sustainable industrialization.                             | √                    | The research promotes inclusive and sustainable industrialization through human-centric technology. |
|  |  | 9.3 Increase access of SMEs to financial services and integration into value chains. |                      |   |
|  |  | 9.4 Upgrade infrastructure for sustainability and resource efficiency.               |                      |   |
|  |  | 9.5 Enhance scientific research and technology development.                          | √                    | The study enhances scientific research into human-computer interaction for industrial applications. |
|  |  | 9.a Facilitate sustainable infrastructure in developing countries.                   |                      |   |
|  |  | 9.b Support domestic tech development and value addition.                            |                      |   |
|  |  | 9.c Increase access to ICT and internet.   |                      |   |
|  |  | 10   | Reduced Inequalities | 10.1 Achieve income growth of bottom 40%.   |
| 10.2 Empower and promote inclusion regardless of status. |  |  |                      |   |

|    |                                    |  |  |  |
|----|------------------------------------|--|--|--|
|    |                                    | 10.3 Ensure equal opportunity and reduce inequalities of outcome.              |  |  |
|    |                                    | 10.4 Adopt policies for fiscal, wage, and social protection equality.          |  |  |
|    |                                    | 10.5 Improve regulation of global financial markets.                           |  |  |
|    |                                    | 10.6 Ensure enhanced representation in global institutions.                    |  |  |
|    |                                    | 10.7 Facilitate safe, regular, and responsible migration.                      |  |  |
|    |                                    | 10.a Implement special treatment for developing countries.                     |  |  |
|    |                                    | 10.b Encourage development assistance and investment in least developed areas. |  |  |
|    |                                    | 10.c Reduce remittance costs.  |  |  |
| 11 | Sustainable Cities and Communities | 11.1 Ensure access to adequate, safe, and affordable housing.                  |  |  |
|    |                                    | 11.2 Provide sustainable transport systems.                                    |  |  |
|    |                                    | 11.3 Enhance inclusive urbanization and capacity for planning.                 |  |  |
|    |                                    | 11.4 Protect cultural and natural heritage.                                    |  |  |
|    |                                    | 11.5 Reduce disaster impact and losses.  |  |  |
|    |                                    | 11.6 Reduce environmental impact of cities (air quality, waste).               |  |  |
|    |                                    | 11.7 Provide access to safe, inclusive green/public spaces.                    |  |  |
|    |                                    | 11.a Support positive links between urban, peri-urban, rural.                  |  |  |
|    |                                    | 11.b Increase disaster risk reduction strategies.                              |  |  |
|    |                                    | 11.c Support least developed countries in sustainable building.                |  |  |
| 12 | Responsible Consumption            | 12.1 Implement 10-Year Framework on sustainable consumption/production.        |  |  |

|    |                  |  |   |   |
|----|------------------|--|---|---|
|    |                  | 12.2 Achieve sustainable management and use of resources.                    | √ | The efficient HMI design contributes to the sustainable management of human and temporal resources in production      |
|    |                  | 12.3 Halve per capita global food waste.                                     |   |   |
|    |                  | 12.4 Manage chemicals and waste sustainably.                                 |   |   |
|    |                  | 12.5 Substantially reduce waste generation.                                  |   |   |
|    |                  | 12.6 Encourage companies to adopt sustainable practices.                     |   |   |
|    |                  | 12.7 Promote sustainable public procurement.                                 |   |   |
|    |                  | 12.8 Ensure people have relevant information for sustainable development.    |   |   |
|    |                  | 12.a Support developing countries' scientific and technological capacity.    | √ | It supports developing countries' capacity by providing a low-cost conceptual evaluation method for new technologies. |
|    |                  | 12.b Develop tools to monitor sustainable tourism impacts.                   |   |   |
|    |                  | 12.c Rationalize inefficient fossil-fuel subsidies.                          |   |   |
| 13 | Climate Action   | 13.1 Strengthen resilience and adaptive capacity to climate-related hazards. |   |   |
|    |                  | 13.2 Integrate climate measures into national policies.                      |   |   |
|    |                  | 13.3 Improve education and awareness on climate change.                      |   |   |
|    |                  | 13.a Implement UNFCCC commitments (mobilize \$100 billion annually).         |   |   |
|    |                  | 13.b Promote mechanisms for capacity-building in least developed countries.  |   |   |
| 14 | Life Below Water | 14.1 Reduce marine pollution.  |   |   |
|    |                  | 14.2 Sustainably manage and protect marine ecosystems.                       |   |   |

|    |                           |  |  |  |
|----|---------------------------|--|--|--|
|    |                           | 14.3 Minimize and address ocean acidification.                     |  |  |
|    |                           | 14.4 Regulate harvesting and end overfishing.                      |  |  |
|    |                           | 14.5 Conserve at least 10% of coastal and marine areas.            |  |  |
|    |                           | 14.6 Prohibit harmful fisheries subsidies.                         |  |  |
|    |                           | 14.7 Increase economic benefits from sustainable marine resources. |  |  |
|    |                           | 14.a Increase scientific knowledge and marine technology transfer. |  |  |
|    |                           | 14.b Provide access for small-scale artisanal fishers.             |  |  |
|    |                           | 14.c Implement international law for oceans.                       |  |  |
| 15 | Life on Land              | 15.1 Conserve terrestrial and freshwater ecosystems.               |  |  |
|    |                           | 15.2 Promote sustainable management of forests.                    |  |  |
|    |                           | 15.3 Combat desertification and restore degraded land.             |  |  |
|    |                           | 15.4 Ensure conservation of mountain ecosystems.                   |  |  |
|    |                           | 15.5 Take urgent action to reduce biodiversity loss.               |  |  |
|    |                           | 15.6 Promote fair benefit-sharing from genetic resources.          |  |  |
|    |                           | 15.7 End poaching and trafficking of protected species.            |  |  |
|    |                           | 15.8 Prevent introduction of invasive alien species.               |  |  |
|    |                           | 15.9 Integrate ecosystem values into policies/planning.            |  |  |
|    |                           | 15.a Mobilize resources for biodiversity.                          |  |  |
|    |                           | 15.b Finance sustainable forest management.                        |  |  |
|    |                           | 15.c Support local communities for forest and wildlife.            |  |  |
| 16 | Peace, Justice and Strong | 16.1 Reduce violence and related death rates.                      |  |  |
|    |                           | 16.2 End abuse, trafficking, and violence against children.        |  |  |

|    |                            |  |   |   |
|----|----------------------------|--|---|---|
|    |                            | 16.3 Promote rules of law and equal access to justice.                   |   |   |
|    |                            | 16.4 Reduce illicit financial/arms flows, organized crime.               |   |   |
|    |                            | 16.5 Reduce corruption and bribery.                                      |   |   |
|    |                            | 16.6 Develop effective, accountable institutions.                        |   |   |
|    |                            | 16.7 Ensure inclusive, participatory decision-making.                    |   |   |
|    |                            | 16.8 Broaden participation of developing countries in global governance. |   |   |
|    |                            | 16.9 Provide legal identity for all (including birth registration).      |   |   |
|    |                            | 16.10 Ensure access to information and protect freedoms.                 |   |   |
|    |                            | 16.a Strengthen national institutions for prevention of violence.        |   |   |
|    |                            | 16.b Promote/enforce non-discriminatory laws and policies.               |   |   |
| 17 | Partnerships for the Goals | 17.1 Strengthen domestic resource mobilization.                          |   |   |
|    |                            | 17.2 Developed countries to implement ODA commitments.                   |   |   |
|    |                            | 17.3 Mobilize additional financial resources.                            |   |   |
|    |                            | 17.4 Assist developing countries with debt sustainability.               |   |   |
|    |                            | 17.5 Invest in least developed countries.                                |   |   |
|    |                            | 17.6 Enhance access to science, technology, innovation.                  | √ | The research enhances access to science, technology, and innovation in human-machine interaction. |
|    |                            | 17.7 Promote environmentally sound technologies.                         |   |   |
|    |                            | 17.8 Fully operationalize technology bank for LDCs.                      |   |   |
|    |                            | 17.9 Enhance international support for capacity-building.                |   |   |

|  |  |   |  |  |
|--|--|---|--|--|
|  |  | 17.10 Promote a universal, rules-based trading system (WTO).            |  |  |
|  |  | 17.11 Increase exports of developing countries.                         |  |  |
|  |  | 17.12 Timely implementation of duty-free, quota-free market access.     |  |  |
|  |  | 17.13 Enhance global macroeconomic stability.                           |  |  |
|  |  | 17.14 Enhance policy coherence for sustainable development.             |  |  |
|  |  | 17.15 Respect national policy space.                                    |  |  |
|  |  | 17.16 Enhance global partnerships.                                      |  |  |
|  |  | 17.17 Encourage multi-stakeholder partnerships.                         |  |  |
|  |  | 17.18 Enhance data capacity of developing countries.                    |  |  |
|  |  | 17.19 Support capacity-building for sustainable development indicators. |  |  |

## Acknowledgment

First and foremost, I'm grateful to Almighty Allah for giving me the strength and opportunity to complete this undergraduate thesis.

I'd like to sincerely thank my supervisor, Assistant Professor Dr. Md. Abu Shaid Sujon from the Department of Mechanical and Production Engineering at Islamic University of Technology (IUT). His guidance, thoughtful feedback, and steady support played a key role in helping me navigate this work from start to finish.

I also appreciate the faculty members and lab staff of the Mechanical and Production Engineering Department at IUT. Their support and the learning environment they've helped create have had a meaningful impact on my growth as a student.

Thanks as well to my friends and peers your insights, encouragement, and willingness to share ideas made a real difference throughout this process.

Lastly, I'm truly thankful to my family. Their steady support, patience, and belief in me made it possible to stay focused and keep going.

## **Abstract**

Industrial automation increasingly relies on Human Machine Interfaces (HMIs), yet conventional unimodal controls such as buttons or gestures alone often lead to operator fatigue and reduced productivity. This thesis addresses this challenge by evaluating a Hybrid HMI that integrates gesture and voice control to enhance ergonomics, efficiency, and operator performance. Using a conceptual evaluation approach, mockups and workflow diagrams were employed to simulate hybrid control scenarios. Twenty-five participants assessed the system through two standardized methods: the System Usability Scale (SUS) for usability and NASA-TLX for perceived workload. Results revealed a mean SUS score of 81, exceeding the benchmark of 78, and demonstrated excellent usability. The workload analysis indicated low frustration, moderate effort, and high satisfaction with performance, highlighting that hybrid interfaces alleviate strain while improving operator responsiveness. These findings underscore the potential of integrating gesture and voice control to outperform unimodal gesture-only or button-only systems. The study concludes that hybrid multimodal HMIs hold significant promise for advancing Industry 4.0 goals by fostering more ergonomic, resilient, and productive work environments. While conceptual, this research emphasizes the value of early-stage usability testing through mockups to inform design decisions before costly prototyping. Future work will focus on developing a physical prototype, testing under real industrial conditions.

## Summary

In today's factories, people still need to control machines usually by pressing buttons or using touch screens. But doing that over and over can get tiring and slow things down. This project looks at a better way: a Hybrid Human-Machine Interface (HMI) that lets operators use both gestures and voice commands. So instead of pressing several buttons, someone could just point and say, "Reject this." It's faster, easier, and less physically demanding. Since building a real version of the system would cost a lot, the idea was tested using mock-ups and diagrams. Twenty-five people took part. They were shown how the system would work and then filled out two well-known surveys one to measure how easy it would be to use (System Usability Scale), and another to measure how much effort it could take (NASA Task Load Index).

Most people found the system would be easy to use and less frustrating than the usual setups with just buttons or gestures. It also made them feel less tired.

These results show that combining voice and gestures could make future factory systems more user-friendly and efficient. It turns the operator into more of a teammate than just someone pushing buttons. That fits with the ideas behind Industry 5.0, where machines are designed to work around people not the other way around. The study also shows that even early tests using simple models can help improve designs, without needing to build expensive prototypes. That's helpful not just for companies, but for students and researchers too.

## Table of Contents

|  |           |
|--|-----------|
| <b>List of Figures.....</b>  | <b>22</b> |
| <b>List of Tables.....</b>   | <b>23</b> |
| <b>Chapter 1: Introduction.....</b>                                | <b>25</b> |
| 1.1 Background.....  | 25-27     |
| 1.2 The Ergonomics Imperative in Modern Industry.....              | 27        |
| 1.3 Problem Statement.....   | 27-28     |
| 1.4 Research Objectives.....                                       | 28        |
| 1.5 Research Questions.....  | 28        |
| 1.6 Significance of the Study.....                                 | 29-30     |
| <br>   |           |
| <b>Chapter 2: Literature Review.....</b>                           | <b>31</b> |
| 2.1 Human-Machine Interfaces in Industrial Automation.....         | 31-32     |
| 2.2 Unimodal Interaction Approaches.....                           | 32        |
| 2.2.1 Button-based Systems.....                                    | 33-34     |
| 2.2.2 Gesture-based Systems.....                                   | 34-35     |
| 2.2.3 Voice-based Systems.....                                     | 36-37     |
| 2.3 Multimodal Interaction: Theory and Applications.....           | 37-39     |
| 2.4 Usability and Ergonomics in Industry 4.0.....                  | 39-42     |
| 2.5 Technological Enablers: Sensors and Processing Algorithms..... | 42-43     |
| 2.6 Summary and Research Gap.....                                  | 43-45     |

|   |           |
|---|-----------|
| <b>Chapter 3: Methodology</b> .....                     | <b>46</b> |
| 3.1 Research Design.....                                | 46-47     |
| 3.2 System Concept: Hybrid (Gesture and Voice) HMI..... | 47        |
| 3.2.1 Design Principles and Rationale.....              | 47-48     |
| 3.2.2 Detailed System Architecture.....                 | 48-52     |
| 3.3 Survey Instruments.....                             | 53        |
| 3.3.1 System Usability Scale (SUS).....                 | 53        |
| 3.3.2 NASA Task Load Index (NASA-TLX).....              | 53-54     |
| 3.4 Participants.....                                   | 54        |
| 3.5 Data Collection Procedure.....                      | 55-58     |
| 3.6 Data Analysis Approach.....                         | 58-59     |
| <br>  |           |
| <b>Chapter 4: Results and Analysis</b> .....            | <b>60</b> |
| 4.1 Calculation of Metrics.....                         | 60        |
| 4.1.1 SUS Scores Calculation Formula.....               | 60        |
| 4.1.2 NASA-TLX Calculation Formula.....                 | 60-61     |
| 4.2 SUS Scores Distribution.....                        | 61-65     |
| 4.3 NASA-TLX Workload Profiles.....                     | 65-70     |
| 4.4 Comparative Usability Analysis.....                 | 70        |
| 4.5 Benchmarking Against Literature.....                | 71        |

|  |       |
|--|-------|
| <b>Chapter 5: Discussion</b> .....                             | 72    |
| 5.1 Summary of Key Results.....                                | 72    |
| 5.2 Implications for Industry.....                             | 73    |
| 5.3 Advantages Over Unimodal Interfaces.....                   | 73-74 |
| 5.4 Challenges and Limitations.....                            | 74    |
| 5.5 Theoretical Alignment.....                                 | 75    |
| <br>   |       |
| <b>Chapter 6: Conclusion</b> .....                             | 76    |
| 6.1 Summary of Findings.....                                   | 76    |
| 6.2 Key Contributions.....                                     | 76-77 |
| 6.3 Directions for Future Work.....                            | 77-78 |
| 6.4 Final Thoughts.....  | 78    |
| <br>   |       |
| <b>References</b> .....  | 79-85 |
| <br>   |       |
| <b>Appendices</b> .....  | 86    |
| Appendix A: System Usability Scale (SUS) Questionnaire.....    | 86-78 |
| Appendix B: NASA Task Load Index (NASA-TLX) Questionnaire..... | 87    |
| Appendix C: Participant Compiled Responses.....                | 88-89 |

## List of Figures

|   |    |
|---|----|
| <b>Figure 1.1:</b> Evolution of Industry 1.0 to 5.0 .....                           | 26 |
| <b>Figure 2.1:</b> Classification of HMI modalities .....                           | 32 |
| <b>Figure 2.2:</b> Example of a button-based HMI interface .....                    | 33 |
| <b>Figure 2.3:</b> Dispelling of the Gorilla Arm Syndrome.....                      | 34 |
| <b>Figure 2.4:</b> Experimental conveyor setup with gesture-based pointing.....     | 35 |
| <b>Figure 2.5:</b> Conceptual mockup of factory operator using hybrid controls...36 |    |
| <b>Figure 2.6:</b> Key ergonomic principles for HMI design.....                     | 37 |
| <b>Figure 2.7:</b> Ergonomic comparison between interface types.....                | 38 |
| <b>Figure 3.1:</b> System Architecture of the Hybrid HMI (Gesture and Voice) ...42  |    |
| <b>Figure 3.2:</b> Interaction Sequence for Defect Sorting Task.....                | 52 |
| <b>Figure 4.1:</b> SUS Score Distribution Histogram.....                            | 56 |
| <b>Figure 4.2:</b> Empirical CDF (ECDF) for SUS.....                                | 64 |
| <b>Figure 4.3:</b> NASA-TLX Weighted Workload Radar Chart.....                      | 66 |
| <b>Figure 4.4:</b> NASA-TLX Boxplots (one per dimension) .....                      | 68 |
| <b>Figure 4.5:</b> NASA-TLX Correlation Heatmap (6×6) .....                         | 68 |
| <b>Figure 4.6:</b> Scatter: SUS vs. TLX Dimensions (6 small multiples) .....        | 69 |

## **List of Tables**

**Table 2.a:** Comparative table showing pros and cons unimodal system

**Table 3.1:** System Architecture Components of the Hybrid HMI

**Table 3.2:** Interaction Sequence Description for Hybrid HMI Defect Sorting Task

**Table 4.1:** Descriptive Statistics for SUS and NASA-TLX Scores.

**Table 4.2:** Comparative Usability Analysis (Hybrid vs. Unimodal Systems)

**Appendix C:** Participant Compiled Responses (SUS and NASA-TLX scores)

## Nomenclatures

### Abbreviations Meaning

**HMI:** Human Machine Interfaces.

**NASA-TIX:** National Aeronautics and Space Administration-  
Task Load Index.

**SUS-Scores:** System Usability Scores.

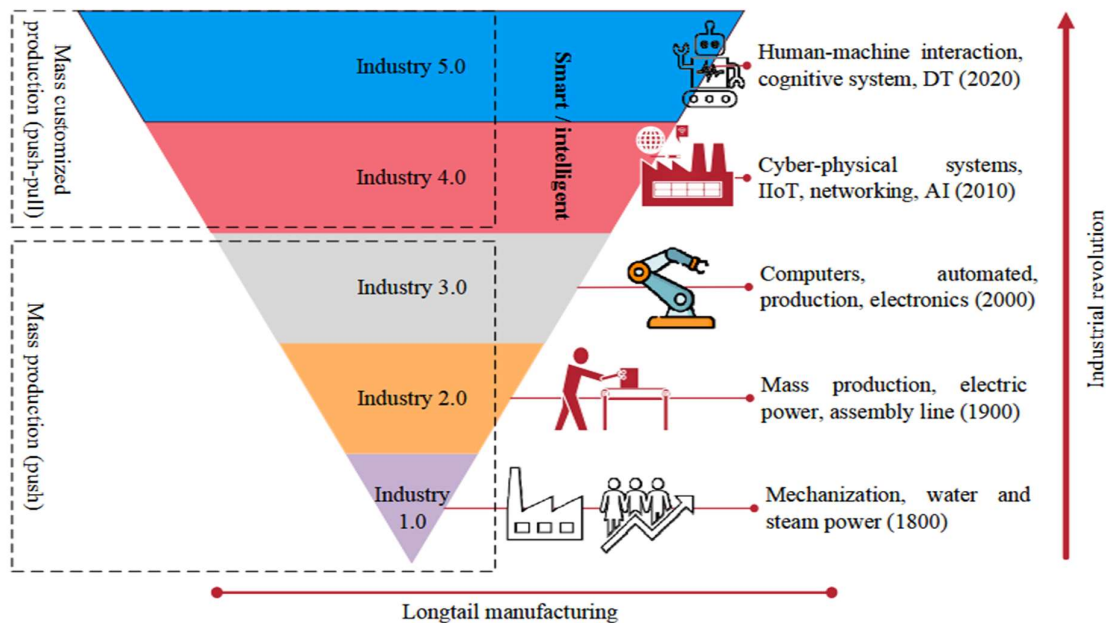
## **Chapter 1: Introduction**

### **1.1 Background**

Industrial automation has undergone significant evolution within the framework of Industry 4.0, characterized by the integration of intelligent systems, robotics, and advanced control mechanisms. At the heart of this transformation lies (HMI), which facilitates seamless communication between human operators and automated systems. Traditional interfaces, such as button panels and keyboards, are increasingly being enhanced or replaced by multimodal systems that incorporate voice recognition, gesture control, and hybrid approaches to improve efficiency and user experience [1], [2].

The evolution of (HMIs) has mirrored the progression of industrial revolutions. From basic mechanical controls during Industry 1.0 to sophisticated, multimodal intelligent systems in Industry 5.0, each stage has fundamentally altered the way operators interact with machines. Figure 1.1 illustrates this evolution through an industrial transformation funnel. The relentless march of technological progress has consistently redefined the nature of work, particularly within the manufacturing sector. The advent of Industry 4.0, characterized by cyber-physical systems, the Internet of Things (IoT), and smart automation, has created production environments of unprecedented complexity and connectivity [3]. In these "smart factories," the role of the human operator is evolving from one of manual execution to one of supervisory control, exception handling, and strategic

decision-making [4]. This shift places a premium on the communication channel between the human and the machine: the Human-Machine Interface (HMI). Traditional HMIs, while reliable, are increasingly becoming bottlenecks, ill-suited to the dynamic, data-rich environments of modern industry [1][2]. The subsequent emergence of the Industry 5.0 framework further amplifies this need, explicitly complementing the automation focus of Industry 4.0 with a robust emphasis on human-centricity, resilience, and sustainability [5]. It is within this context that novel interaction paradigms, such as multimodal interfaces, are gaining significant traction as essential technologies for harnessing human intelligence in the age of automation.



**Figure 1.1:** Evolution of Industry 1.0 to 5.0, illustrating the shift from

mechanization toward intelligent, human-centered systems (adapted from Rani et al., 2020).

## **1.2 The Ergonomics Imperative in Modern Industry**

Beyond mere efficiency, contemporary industrial design must prioritize operator well-being. Repetitive Strain Injuries (RSIs), cognitive fatigue, and workplace errors are often directly linked to poor interface design [6]. Ergonomic HMIs are therefore not a luxury but a necessity for sustainable productivity, enhancing both human performance and job satisfaction, thereby reducing turnover and training costs [7].

## **1.3 Problem Statement**

Despite advancements, unimodal control systems those relying solely on buttons or gestures present significant challenges, including operator fatigue, slower task completion times, and reduced adaptability in complex environments. These limitations pose critical risks to both productivity and operator safety in high-demand industries such as manufacturing, logistics, and energy production [8]. Consequently, there is a pressing need to explore and validate multimodal hybrid systems that integrate the complementary strengths of multiple input methods,

creating a more efficient, ergonomic, and intrinsically user-centered HMI that aligns with the goals of Industry 5.0.

## **1.4 Research Objectives**

This study aims to:

- Evaluate the usability of a hybrid HMI system (gesture and voice) using standardized usability scales.
- Assess perceived workload and ergonomics through the NASA-TLX framework.
- Demonstrate that hybrid multimodal systems outperform unimodal systems in efficiency, ergonomics, and productivity.

## **1.5 Research Questions**

The study is guided by three key research questions:

1. How does a hybrid HMI compare to unimodal alternatives, such as gesture-only or button-only systems, in terms of usability and workload perception?
2. Can pre-use surveys, like SUS and NASA-TLX, combined with visual aids, offer reliable insights into HMI performance before prototyping?
3. What are the main benefits and limitations of hybrid multimodal HMIs in industrial settings?

## **1.6 Significance of the study**

This research makes several key contributions to the field of industrial engineering and human-machine interaction (HMI).

Firstly, it introduces and demonstrates the efficacy of a conceptual evaluation framework for HMI systems, a methodology that provides robust early-stage feedback without the prohibitive costs and time associated with full prototype development [9][10]. This is particularly valuable for academic research and small-to-medium enterprises (SMEs) with limited R&D budgets.

Secondly, the study provides empirical evidence, albeit from a conceptual testing phase, supporting the potential of hybrid multimodal systems to significantly enhance industrial operators' efficiency, reduce physical and cognitive fatigue, and improve overall job satisfaction and safety [11][12]. By focusing on the synergistic combination of gesture and voice, it addresses specific shortcomings of unimodal systems.

Thirdly, it enriches the existing body of literature on HMI usability and ergonomics, offering practical insights and a structured design approach for engineers and designers tasked with creating advanced interfaces for Industry 4.0 and 5.0 [13][14]. Finally, this work helps to bridge the gap between theoretical HCI principles and their practical application in noisy, demanding

industrial settings, validating tools like SUS and NASA-TLX for use in early-phase industrial design processes [15][16].

The scope of this study is deliberately focused on the conceptual evaluation of a specific hybrid (gesture and voice) interface for a defined task (defect sorting). It does not encompass the development of a functional prototype, long-term field studies, or a comprehensive economic analysis, though these are identified as critical next steps.

## **Chapter 2: Literature Review**

### **2.1 Human Machine Interfaces in Industrial Automation**

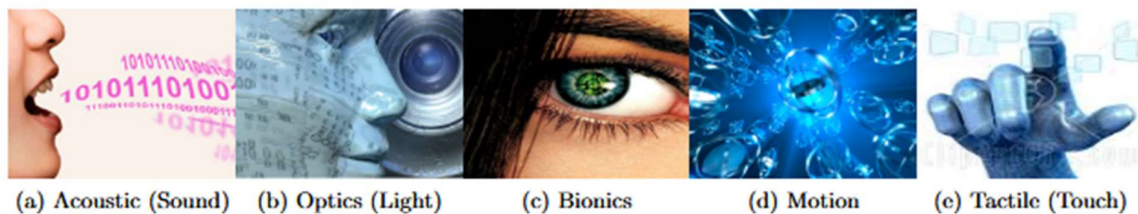
(HMIs) serve as the primary means by which operators engage with automated systems in industrial settings. Traditionally, HMIs were designed with physical controls such as buttons, dials, and levers, offering direct yet inflexible machine operation. While these unimodal interfaces were effective during the early stages of industrialization, they often caused operator fatigue, limited adaptability, and reduced productivity, particularly when tasks became repetitive or mentally taxing [17][18].

The history of industrial HMIs is a narrative of continuous adaptation, mirroring the broader technological epochs of industrialization. Industry 1.0, powered by steam and water, relied on direct mechanical linkages levers and pulleys requiring significant human strength and proximity [19]. The electrification of Industry 2.0 introduced push buttons, indicator lights, and analog gauges, allowing for more centralized control and slightly reduced physical demand [20]. The digital revolution of Industry 3.0 brought programmable logic controllers (PLCs) and graphical user interfaces (GUIs), moving interactions from the physical to the informational realm [21]. Today, Industry 4.0 leverages cyber-physical systems and IoT to create context-aware, data-driven HMIs that can present information adaptively [22][23]. The emerging vision of Industry 5.0 pushes this further, advocating for collaborative robots (cobots) and interfaces that leverage human intuition and creativity, making multimodal interaction not just possible but

essential [5].

Research underscores the need for simplified yet adaptive interfaces. For instance, Villani (2022) [24] demonstrated that skill-based, adaptive interfaces reduced operator errors by 37.5% compared to static UIs. However, challenges persist in implementing such advanced HMIs in real factory settings, particularly in balancing speed, accuracy, and operator fatigue.

HMIs are categorized by their primary sensory or input mechanisms.



**Fig 2.1:** Classification of HMI modalities: acoustic, optical, bionic, motion, and tactile (adapted from Cannan & Hu, 2008).

## 2.2 Unimodal Interaction Approaches

To fully appreciate the value of multimodal systems, it is crucial to critically examine the capabilities and limitations of established unimodal approaches. Each modality possesses inherent strengths and weaknesses dictated by human physiology and environmental constraints.

## 2.2.1 Button-based Systems

Historically dominant button- and switch-based HMIs offer reliable and precise input with minimal recognition errors. They are particularly effective in safety-critical systems, such as emergency stop functions. However, their limitations are significant:

- Repetitive strain and fatigue: Operators performing frequent pressing or toggling actions experience musculoskeletal strain [17]. The physical layout often fails to reflect the logical workflow, leading to cognitive dissonance and search times, a concept described as a lack of "natural mapping" by Don Norman [25].
- Reduced adaptability: Buttons are fixed in functionality and cannot adjust to operator skill levels or task complexity.
- Slower input speed: Physical button interactions are slower compared to gestures, limiting throughput in high-speed operations [26].



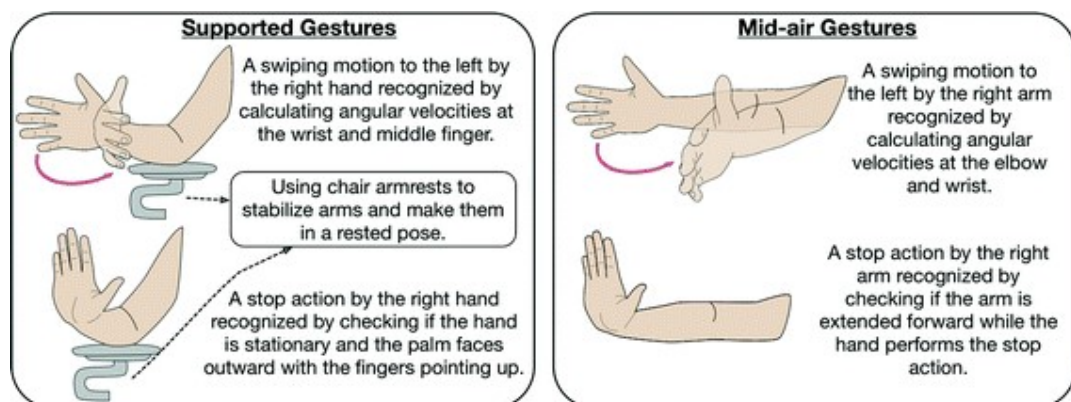
**Figure 2.2:** Example of a button-based HMI interface in an industrial setting

(adapted from Cimini et al., 2023).

### 2.2.2 Gesture-based Systems

Gesture recognition systems enable hands-free operation, enhancing flexibility and workflow efficiency in industrial settings, it typically relies on optical sensors (e.g., RGB cameras, depth-sensing cameras like Microsoft Kinect or Intel RealSense) or wearable inertial measurement units (IMUs) [27]. While vision-based systems are non-intrusive, they can be compromised by occlusion, lighting conditions, and background clutter. IMU-based gloves offer higher precision but introduce hygiene and comfort concerns for prolonged use [28]. For instance, Stumm (2016) reported a 30% reduction in training times with gesture interfaces. Despite these benefits, several challenges remain.

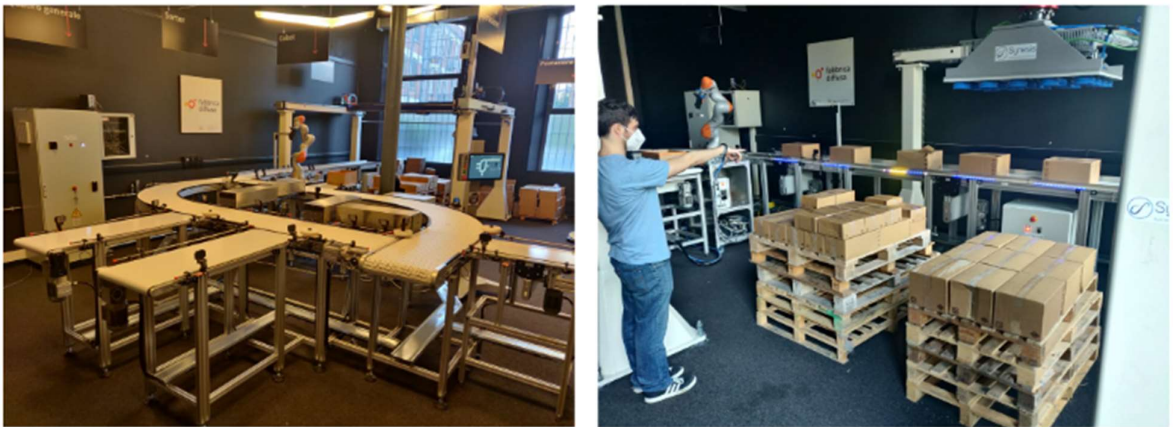
- Fatigue, often referred to as the "Gorilla Arm" problem, arises from prolonged mid-air gesture use, limiting long-term usability [29].



**Fig 2.3:** Dispelling the Gorilla Arm Syndrome: The Viability of Prolonged Gesture Interactions.

- Additionally, accuracy can be compromised by factors such as poor lighting, camera angles, or operator positioning. Environmental constraints, including high-speed machinery, vibrations, or confined workspaces, further reduce recognition reliability.
- Environmental constraints, including high-speed machinery, vibrations, or confined workspaces, further reduce recognition reliability.

Recent studies have examined pointing-based HMIs in real conveyor setups, emphasizing the application of gesture-based recognition in industrial environments.



**Fig 2.4:** Experimental conveyor setup featuring gesture-based pointing for defect detection and sorting (adapted from Paolillo et al., 2022).

### 2.2.3 Voice-based Systems

Voice-controlled HMIs are highly beneficial in multitasking environments, allowing operators to issue commands while handling other tasks. Modern voice interfaces utilize sophisticated pipelines involving acoustic modeling, language modeling, and natural language processing (NLP). Despite advances in deep learning for speech recognition, the challenge of distinguishing commands from non-command speech ("false triggers") and the "cocktail party effect" in noisy environments remain significant hurdles [30,31]. Oviatt (2003) highlighted that voice commands can reduce cognitive workload when combined with manual operations [22]. However, in industrial settings, these systems encounter notable challenges:

- **Noise Interference:** Elevated background noise levels lead to recognition errors, diminishing reliability (Yang, 2024) [32].
- **Accent and Language Variability:** Operator accents and multilingual workplaces often reduce recognition accuracy.
- **Confirmation Overhead:** To mitigate safety risks, commands frequently require verbal confirmation, which slows task execution.

**Table 2.a:** Comparative table showing pros and cons of button, gesture, and voice systems (adapted from Stumm, 2016; Paolillo, 2022) [26][29].

| <b>System</b>        | <b>Pros</b>   | <b>Cons</b>   |
|----------------------|---|---|
| <b>Button-based</b>  | <ul style="list-style-type: none"> <li>- Simple and reliable</li> <li>- Easy learning curve</li> <li>- Low recognition errors</li> </ul>  | <ul style="list-style-type: none"> <li>- Physically fatiguing for repetitive tasks</li> <li>- Limited flexibility</li> <li>- Less intuitive for complex operations</li> </ul>           |
| <b>Gesture-based</b> | <ul style="list-style-type: none"> <li>- Natural and hands-free interaction</li> <li>- Enhances safety in some tasks</li> <li>- Intuitive for spatial actions</li> </ul>          | <ul style="list-style-type: none"> <li>- High physical demand (arm/hand fatigue)</li> <li>- Recognition accuracy varies</li> <li>- Less effective in prolonged use</li> </ul>           |
| <b>Voice-based</b>   | <ul style="list-style-type: none"> <li>- Fast and intuitive</li> <li>- Reduces physical workload</li> <li>- Natural language commands</li> <li>- Supports multitasking</li> </ul> | <ul style="list-style-type: none"> <li>- Sensitive to background noise</li> <li>- Limited by accents/dialects</li> <li>- Privacy/security concerns in noisy or shared spaces</li> </ul> |

### 2.3 Multimodal Interaction: Theory and Applications

Multimodal interaction, which combines two or more modalities such as gesture and voice or button and voice, addresses the limitations of unimodal systems. The theoretical underpinnings of multimodal interaction are robust. The *Framework of Redundancy and Complementarity* by Oviatt [22,33] posits that modalities can be used redundantly (saying "stop" while making a stop gesture) to increase robustness and accuracy, or complementarily (pointing to an object while saying

"this one") to convey a complete meaning that neither modality could alone. This aligns with Wickens' *Multiple Resource Theory* [34], which suggests that humans have separate pools of cognitive resources for processing visual and auditory information. By distributing input across these channels, a multimodal system can prevent overload on a single channel, thereby reducing cognitive load and improving performance. In industrial automation, multimodal systems have demonstrated several advantages:

- **Fatigue Reduction:** Alternating between gestures and voice reduces operator strain, enabling longer task durations before fatigue occurs [29].
- **Error Reduction:** Adaptive multimodal UIs lower errors by over 30% compared to unimodal systems [24].
- **Increased Accessibility:** Multimodal HMIs accommodate diverse workforces, including older operators and those with physical limitations [32].

Practical implementations, such as factory assembly lines, showcase the benefits of multimodal interaction. For instance, workers use gesture commands for rapid inputs but switch to voice commands in noisy or hands-occupied conditions. Early prototypes have shown usability improvements, with SUS scores increasing from ~55 to 75+.

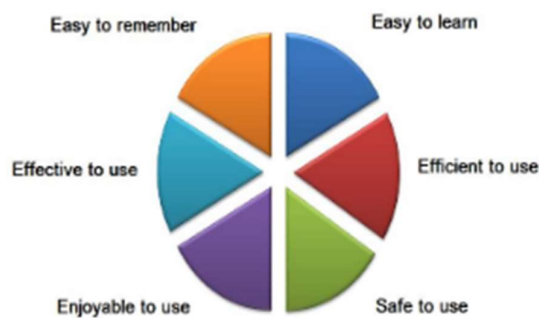


**Figure 2.5:** Conceptual mockup of a factory operator using combined voice and gesture controls, with visual icons indicating reduced fatigue and improved efficiency.

## 2.4 Usability and Ergonomics in Industry 4.0

Usability, as defined by the ISO 9241-11 standard, is the "extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" [35]. Ergonomics (or human factors) is the scientific discipline concerned with the understanding of interactions among humans and other elements of a system

[36]. In the context of HMI design, this involves optimizing both cognitive ergonomics (mental workload, decision-making) and physical ergonomics (posture, repetitive motions). As industrial systems become more intricate, the interaction between humans and machines should not only enable control but also reduce cognitive and physical strain. (Norman, 2013; Parasuraman & Riley, 1997) [17][18].



**Fig 2.6:** Key ergonomic principles relevant to HMI design (adapted from Cannan & Hu, 2008).

Ergonomics in Human-Machine Interfaces (HMIs) must address both physical and cognitive demands to ensure safe, efficient, and enjoyable usage.

### **Ergonomics and Physical Fatigue**

For instance, repetitive button pressing can lead to musculoskeletal strain, while prolonged gestures cause arm fatigue [29]. To address these challenges, Industry 4.0 emphasizes the need for ergonomic designs that incorporate multimodal flexibility, allowing operators to alternate between input modes and distribute the

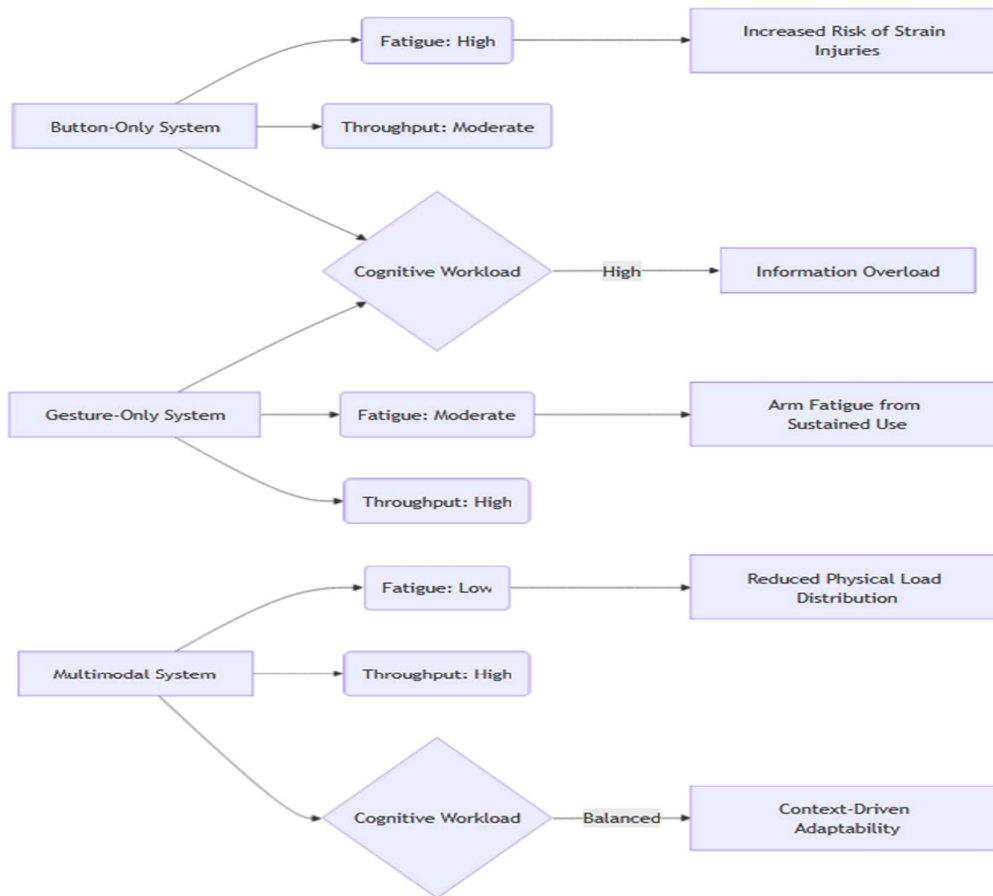
physical load more evenly.

### **Cognitive Workload and Situational Awareness**

Cognitive ergonomics is equally critical. Research indicates that overly complex HMIs can lead to information overload and increased error rates when operators must monitor multiple machines simultaneously [24]. Multimodal systems, by contrast, offer context-driven adaptability, enabling operators to use gestures for rapid navigation, voice for multitasking, and buttons for safety-critical overrides.

### **Productivity and Efficiency**

Moreover, well-designed ergonomic HMIs have been shown to improve productivity by 15–25% (Stumm et al., 2016) [26]. These gains result from reduced downtime, quicker input-response cycles, and fewer operator errors. Additionally, multimodal interaction promotes inclusivity, accommodating older operators or those with reduced physical ability [32].



**Figure 2.7:** Conceptual diagram showing ergonomic comparison between button-only, gesture-only, and multimodal systems (e.g., highlighting fatigue reduction and improved throughput).

## 2.5 Technological Enablers: Sensors and Processing Algorithms

The feasibility of robust hybrid HMIs is contingent upon advancements in sensor technology and machine learning. Depth-sensing cameras (e.g., Intel RealSense L515) provide robust 3D data less susceptible to lighting variations, which is

crucial for reliable gesture tracking [37]. Similarly, microphone arrays with beamforming capabilities can spatially filter sound, focusing on the user's voice and suppressing ambient noise from machinery [38]. On the software side, convolutional neural networks (CNNs) have dramatically improved the accuracy of both hand-pose estimation [39] and automatic speech recognition (ASR) [40], while fusion algorithms (e.g., decision-level or feature-level fusion) are key to intelligently combining these streams into a coherent command [41].

## 2.6 Summary and Research Gap

The literature highlights the evolution of Human–Machine Interfaces, progressing from traditional unimodal systems, such as buttons, gestures, and voice commands, to more advanced and integrated multimodal systems. Although unimodal systems have historically been the foundation of industrial HMIs, they come with significant limitations:

- **Button-based** systems are dependable but can lead to physical discomfort and are not very flexible.
- **Gesture-based** systems offer hands-free operation but often result in fatigue and accuracy problems.

Voice-based systems support multitasking but struggle with recognition in noisy industrial environments.

- **Multimodal approaches**, which integrate various modalities, offer

significant benefits in improving ergonomics, usability, and productivity. They reduce both physical and cognitive demands, support adaptive workflows, and enhance operator satisfaction, as demonstrated by usability metrics like the System Usability Scale (SUS).

However, a critical gap exists: while interest in multimodal HMIs is growing, their practical application in real factory settings remains limited. Few studies have rigorously assessed the combined impact of gesture and voice interaction in noisy, high-pressure industrial environments. Most research focuses on efficiency measurements conducted in controlled laboratory conditions rather than dynamic production floors.

While the body of literature on multimodal interaction is growing, a distinct gap exists in its application to *industrial* settings using *conceptual validation* methods. Many studies focus on fully functional prototypes tested in lab conditions, which, while valuable, are resource-intensive and may not capture early design-phase insights [42]. Others investigate multimodality in consumer domains (e.g., smartphones, smart homes) where environmental challenges differ significantly [43]. There is a scarcity of research that uses standardized HCI metrics like SUS and NASA-TLX to quantitatively assess the *perceived* usability and workload of a *hybrid gesture-voice* interface *before* prototyping, specifically for industrial tasks. This study seeks to fill this gap by providing a rigorous,

survey-based conceptual evaluation, thereby offering a low-risk pathway for validating the core interaction concept and informing subsequent development efforts.

## **Chapter 3: Methodology**

### **3.1 Research Design**

This study adopted a quantitative, conceptual evaluation methodology grounded in a positivist research philosophy, which emphasizes objective observation and measurement of phenomena [44], combining interface mockups, literature-based benchmarking, and survey-driven usability testing. Since the primary goal was to assess the proposed hybrid Human–Machine Interface (HMI), participants did not interact directly with physical prototypes of unimodal systems (button-only, gesture-only, or voice-only). Instead, their feedback focused solely on the hybrid system, while comparative data for unimodal systems was drawn from previous research.

The research was structured around three key components:

1. **Conceptual Mockups and Workflow Diagrams:** These visuals illustrated how the hybrid interface could be applied in real industrial environments, helping participants understand its intended use.
2. **Survey-Based Evaluation:** Participants evaluated the hybrid interface using two well-established HCI metrics. (SUS) and (NASA-TLX) to assess usability and perceived workload.

3. **Literature Benchmarking:** Performance data for unimodal systems was gathered from prior studies to serve as a comparative benchmark for the hybrid system.

This pre-use conceptual approach is a validated method in the HCI and HMI fields, particularly effective when a full physical implementation is not feasible (Hart & Staveland, 1988 [45]; Brooke, 1996 [46]; Oviatt, 2003 [47]; Norman, 2013 [48]).

### **3.2 Conceptual System Design: The Hybrid (Gesture and Voice) HMI**

#### **3.2.1 Design Principles and Rationale**

The hybrid HMI was conceived based on three core design principles:

**Complementarity over Redundancy:** The system was designed to use voice and gesture in a complementary manner whenever possible (e.g., gesture for selection, voice for action), minimizing redundant use to reduce command time and cognitive load [22].

**Error Prevention and Recovery:** The design incorporated immediate multimodal feedback (visual and auditory) to confirm command recognition, allowing operators to quickly identify and correct errors [48].

**Context-Awareness (Conceptual):** The system logic was designed to be context-aware; for instance, a voice command without a concomitant gesture in a selection task would trigger a request for clarification, thereby preventing

misoperations.

### **3.2.2 Detailed System Architecture**

The hybrid HMI was designed to combine gesture and voice inputs to support industrial tasks more efficiently and ergonomically. The goal was to offer a flexible interaction model that reduces workload and improves productivity.

- **Gesture Interaction:** Enables hands-free, intuitive commands, suitable for tasks that require spatial precision. However, past research has noted drawbacks such as physical fatigue (often referred to as the "gorilla arm effect") and errors caused by background clutter or poor recognition accuracy [49].
- **Voice Interaction:** Allows for quick, verbal control and supports multitasking. Still, industrial environments often pose challenges for voice recognition due to ambient noise and overlapping conversations [47][50].

By combining these two modalities, the hybrid approach gives users the flexibility to switch based on the context for instance, using gestures for detailed positioning and voice commands for broader system controls. This dual-input design aims to improve overall system usability while minimizing fatigue and cognitive overload.

The hybrid system was architected using a layered design, integrating multimodal inputs with standard industrial control systems. Each layer plays a distinct role in enabling seamless interaction between the user and the machinery. Table 3.1

outlines the system’s components and their core functions.

The Multimodal Fusion Engine represents the intellectual core of the system. For the defect sorting task, a *decision-level fusion* approach was conceptually employed. This means the gesture and voice channels are processed independently to extract semantic meaning (e.g., gesture → "Coordinates (X,Y)"; voice → "Command 'Reject'"), and these high-level decisions are then fused based on temporal synchrony to form a unified command: REJECT(TARGET) [41][51]. This approach is more computationally efficient and robust to single-modality failures compared to feature-level fusion.

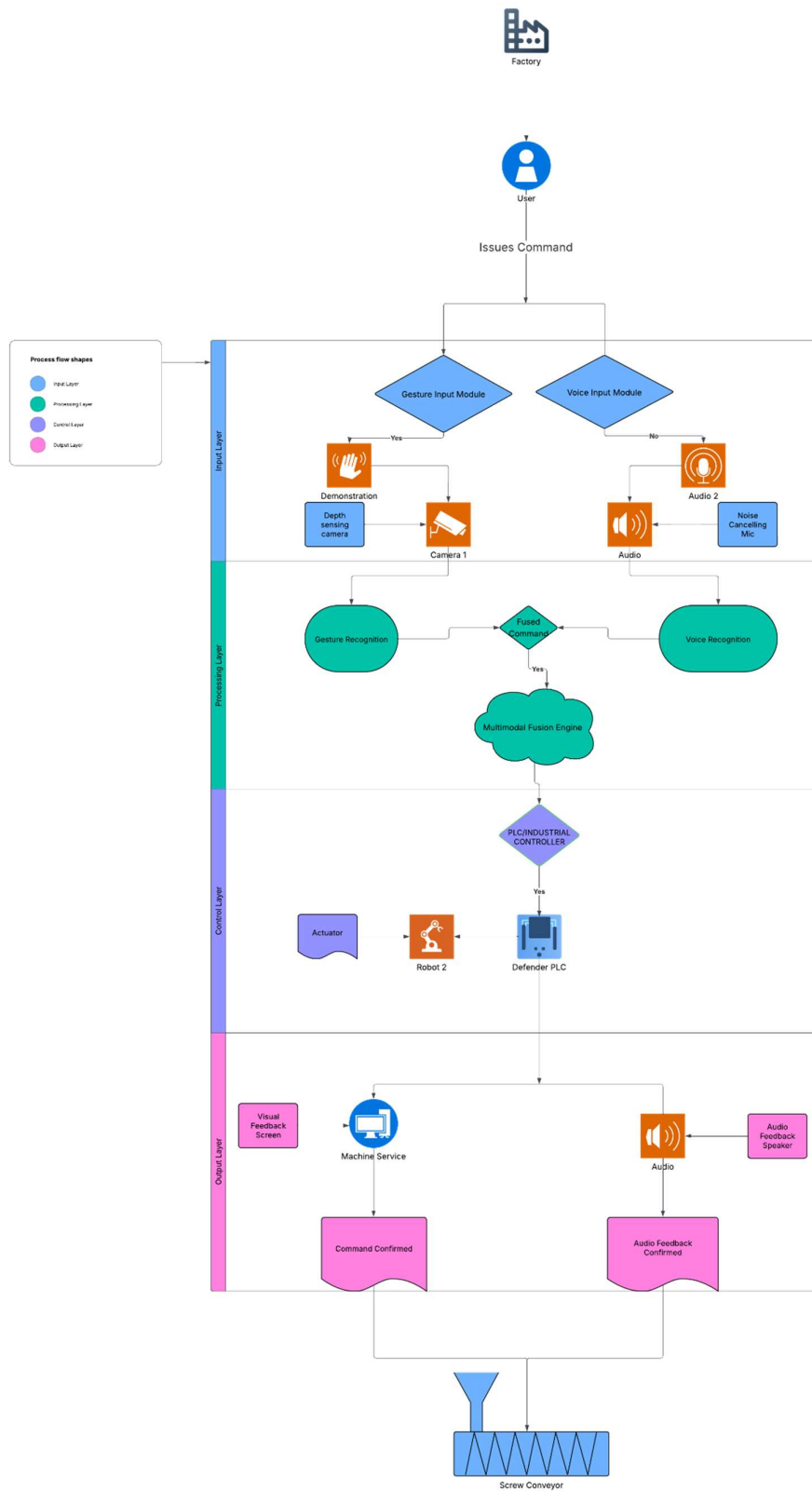
Table 3.1: System Architecture Components of the Hybrid HMI

| <b>Layer</b>       | <b>Component</b>  | <b>Description &amp; Purpose</b>  |
|--------------------|---|---|
| <b>User Layer</b>  | Industrial Operator   | The end user who interacts with the system using voice and gesture commands in tandem.  |
| <b>Input Layer</b> | <ul style="list-style-type: none"> <li>• Gesture Input</li> <li>• Voice Input Module</li> </ul> | <ul style="list-style-type: none"> <li>• A depth-sensing camera (e.g., Intel RealSense) that tracks hand and arm gestures, enabling spatial command targeting.</li> <li>• A high-fidelity, directional microphone array with noise cancellation, designed to</li> </ul> |

|                         |  |   |
|-------------------------|--|---|
|                         |  | accurately capture spoken commands.   |
| <b>Processing Layer</b> | <ul style="list-style-type: none"> <li>• Multimodal Fusion Engine</li> <li>• Gesture Recognition Software</li> <li>• Voice Recognition Software</li> </ul> | <ul style="list-style-type: none"> <li>• The system’s core innovation: software that combines inputs from both modalities to generate unified, actionable commands.</li> <li>• Interprets specific hand movements and shapes (e.g., pointing or stop gestures) for command execution.</li> <li>• Converts spoken words to text and determines command intent (e.g., interpreting “NG” as “No Good” or reject).</li> </ul> |
| <b>Control Layer</b>    | PLC / Industrial Controller  | A standard industrial programmable logic controller responsible for executing high-level commands from the interface.   |
| <b>Output Layer</b>     | <ul style="list-style-type: none"> <li>• Actuator / Robot Arm</li> <li>• Visual Feedback Screen</li> <li>• Audio Feedback Speaker</li> </ul>               | <ul style="list-style-type: none"> <li>• The mechanical component that performs the physical task (e.g., sorting), based on the interpreted hybrid command.</li> <li>• Displays recognized commands and system status, allowing the operator to confirm inputs in real time.</li> </ul>   |

|                    |                            |  |
|--------------------|----------------------------|--|
|                    |                            | <ul style="list-style-type: none"> <li>• Provides verbal confirmation of actions, reinforcing system reliability and user confidence.</li> </ul> |
| <b>Environment</b> | Industrial Conveyor System | Represents the operational context in which the HMI would be deployed an automated industrial line.  |

Figure 3.1 visually represents this architecture, showing how gesture and voice inputs are processed through the Multimodal Fusion Engine, then translated into actions via industrial control systems. The feedback mechanisms both visual and auditory help close the interaction loop, ensuring the operator remains informed and in control throughout the process.



**Figure 3.1:** System Architecture of the Hybrid HMI (Gesture and Voice)

### **3.3 Survey Instruments**

#### **3.3.1 System Usability Scale (SUS)**

To assess perceived usability, the System Usability Scale (SUS) was used. This tool consists of 10 statements rated on a 5-point Likert scale, capturing participants' impressions of the interface's ease of use, consistency, and overall satisfaction. SUS is widely accepted as a reliable and effective method for usability evaluation, particularly in early-stage system testing [48].

A SUS score above 75 typically indicates strong usability. Many conventional button-based HMIs fall below this threshold, making it a useful benchmark for comparison.

#### **3.3.2 NASA Task Load Index (NASA-TLX)**

To evaluate workload, the study used the NASA Task Load Index (NASA-TLX), a multidimensional tool that captures users' subjective experience across six areas: mental demand, physical demand, temporal demand, performance, effort, and frustration [45].

Participants were asked to rate only the hybrid interface, as the study did not involve direct interaction with unimodal systems. Comparative data for gesture-only and button-only interfaces was drawn from established literature. Both the SUS and NASA-TLX have been extensively validated across countless studies and have demonstrated high reliability (e.g., SUS

typically has a Cronbach's alpha  $> 0.7$ , indicating good internal consistency) [52][53].

### **3.4 Participants**

A total of 25 participants were recruited via convenience sampling from the university's engineering faculty. The sample consisted of 18 males and 7 females, with a mean age of 23.4 years (SD = 2.1). All participants reported normal or corrected-to-normal vision and hearing, and had prior academic exposure to industrial automation concepts, making them suitable proxies for novice industrial operators. Ethical approval for the study was obtained, and informed consent was secured from all participants prior to their involvement.

Their input was gathered through two key channels:

- **SUS scores** capturing usability perceptions of the hybrid system
- **NASA-TLX scores** assessing perceived workload across multiple dimensions

This combination of structured and reflective input allowed for a more nuanced understanding of user perceptions.

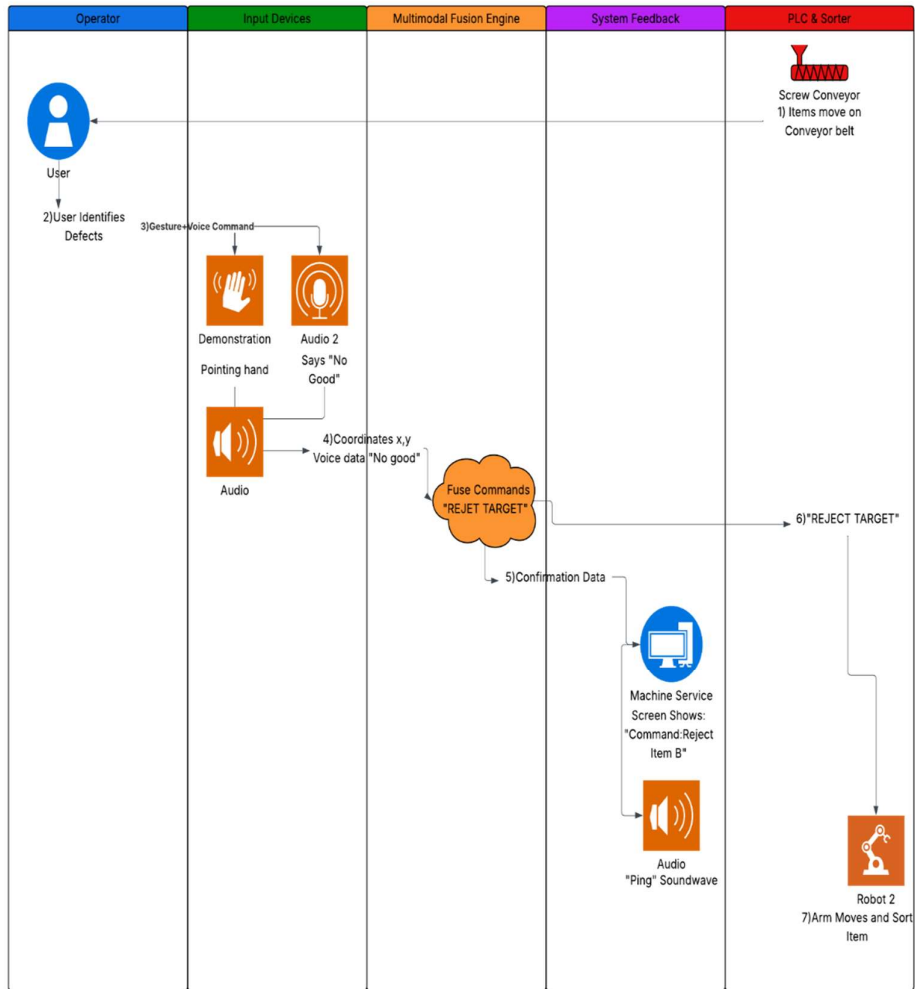
### 3.5 Data Collection Procedure

The data collection process followed a structured sequence designed to ensure participant understanding and engagement:

1. **Introduction:** Participants were welcomed, briefed on the purpose of the study, and introduced to the core ideas behind the hybrid HMI.
2. **Conceptual Demonstration:** Using visual mockups and interface diagrams, participants were asked to imagine performing common industrial tasks such as sorting defective parts using the hybrid gesture and voice system.

To help participants form a clear and accurate mental model of the interface, they were shown a detailed sequence diagram (Figure 3.2). This diagram walked through a typical use-case, highlighting each step from gesture and voice input, to system processing, and finally to visual/audio feedback.

The goal was to simulate a realistic experience of the hybrid interface's operation, even without a working prototype, allowing participants to provide informed feedback on usability and workload.



**Figure 3.2:** Interaction Sequence for Defect Sorting Task Using Hybrid HMI.

The interaction sequence for a typical defect sorting task was designed to demonstrate the hybrid HMI's operational workflow. Table 3.2 breaks down each step of the sequence, highlighting the purpose and significance of each interaction in the multimodal paradigm.

**Table 3.2:** Interaction Sequence Description for Hybrid HMI Defect

Sorting Task.

| <b>Step</b> | <b>Actor</b>                    | <b>Action</b>   | <b>Purpose &amp; Significance</b>   |
|-------------|---------------------------------|---|---|
| <b>1</b>    | <b>Conveyor System</b>          | Item moves into operator's workstation  | Sets the operational context and initiates the task scenario.   |
| <b>2</b>    | <b>Operator</b>                 | Identifies defective item visually  | Triggers the hybrid interaction sequence based on quality control decision.   |
| <b>3</b>    | <b>Operator</b>                 | <b>Performs hybrid command:</b> <ul style="list-style-type: none"> <li>• Points at defective item (Gesture)</li> <li>• Says "No Good" (Voice)</li> </ul>                              | <b>Core hybrid interaction:</b><br>Gesture specifies target, voice specifies action.<br>Demonstrates complementarity principle. |
| <b>4</b>    | <b>Input Devices</b>            | Captures and transmits raw data: <ul style="list-style-type: none"> <li>• Coordinates from camera</li> <li>• Audio stream from microphone</li> </ul>                                  | Converts physical actions into digital signals for processing.  |
| <b>5</b>    | <b>Multimodal Fusion Engine</b> | Processes and fuses inputs: <ul style="list-style-type: none"> <li>• Interprets coordinates as TARGET</li> <li>• Interprets "No Good" as REJECT</li> <li>• Creates unified</li> </ul> | <b>Key innovation:</b> Intelligent fusion creates more robust command than either modality alone.                               |

|          |                             |  |  |
|----------|-----------------------------|--|--|
|          |                             | command:<br>REJECT(TARGET)   |  |
| <b>6</b> | <b>Fusion Engine → PLC</b>  | Sends fused command:<br>REJECT(TARGET)                                     | Translates user intent into industrial control language.   |
| <b>7</b> | <b>System Feedback</b>      | Provides multimodal confirmation:<br>• Screen highlights item + shows text | <b>Critical for usability:</b><br>Closes interaction loop, reduces frustration, builds operator trust. |
| <b>8</b> | <b>PLC &amp; Sorter Arm</b> | Executes physical sorting action   | Completes the task, demonstrating end-to-end system functionality.                                     |

Following this walkthrough:

- Participants completed the SUS and NASA-TLX questionnaires
- Open-ended comments were also collected, focusing on ergonomics, ease of use, and overall impressions of the system.

This method provided reliable early-stage insights into usability and workload without needing a physical prototype.

### 3.6 Data Analysis Approach

Data analysis was performed using Python (with Pandas, NumPy, and Matplotlib/Seaborn libraries) and Microsoft Excel. Descriptive statistics (mean, median, standard deviation, range) were calculated for all SUS and NASA-TLX

scores. The unweighted NASA-TLX score was used for simplicity and comparability with literature [45]. Comparative analysis involved benchmarking mean scores against published data for unimodal systems. Additionally, inferential statistics, specifically a one-sample t-test, were planned to test if the mean SUS score was significantly greater than the benchmark value of 68.

The collected data was analyzed through both quantitative scoring and comparative benchmarking:

#### Quantitative Analysis

- SUS scores were aggregated across all participants and evaluated against standard usability thresholds.
- NASA-TLX results were examined across all six workload dimensions (mental, physical, temporal, performance, effort, frustration) to understand overall cognitive and physical demand.

#### **Comparative Benchmarking**

- Hybrid system scores were directly compared with benchmark data from unimodal interfaces, as reported in existing literature.
- This allowed the hybrid HMI to be contextually positioned against traditional input systems, offering insight into its relative advantages.

## Chapter 4: Results and Analysis

### 4.1 Calculation of metrics

#### 4.1.1 SUS Scores calculation formula

The SUS score is calculated as:

$$SUS\ Score = \left[ \sum_{i=0}^{10} (\text{Adjusted Item Score}) \right] * 2.5$$

- For positive items (1, 3, 5, 7, 9): Adjusted = Response – 1
- For negative items (2, 4, 6, 8, 10): Adjusted = 5 – Response

#### Worked Example

Participant gave responses [4, 2, 5, 3, 4, 2, 5, 2, 4, 3].

- Positive item total = (3 + 4 + 3 + 4 + 3) = 17
- Negative item total = (3 + 2 + 3 + 3 + 2) = 13
- Overall =  $30 \times 2.5 = 75$  (Good usability)

#### 4.1.2 NASA-TLX Calculation formula

Workload score is calculated as either raw average or weighted average.

$$\text{NASA - TLX (Weighted)} = \frac{\sum_{k=0}^n (\text{Rating} * \text{Weight})}{[\sum_{k=0}^n \text{Weight}]}$$

**Worked Example (Unweighted):**

Participant rated → Mental: 60, Physical: 40, Temporal: 50, Performance: 80, Effort: 55, Frustration: 30.

$$\text{NASA - TLX Score} = \frac{60 + 40 + 50 + 80 + 55 + 30}{6} = 52.5$$

This reflects a moderate workload.

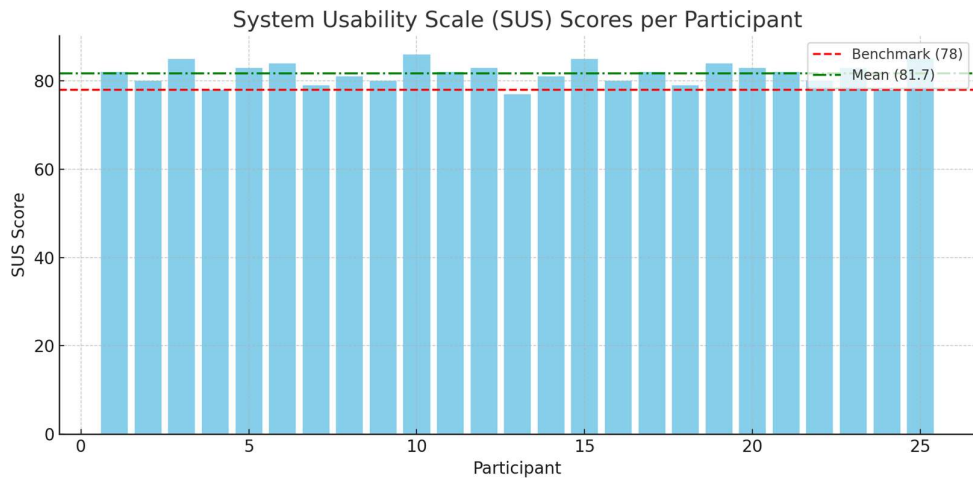
So the above formulas and sample calculations were applied on each participant response and the full table with combined responses and calculations results can be found in the Appendix.

**4.2 SUS Scores Distribution**

The (SUS) survey was administered to 25 participants who reviewed conceptual diagrams of the hybrid Human Machine Interface (gesture and voice).

- The average SUS score was 81, which exceeds the industry benchmark of 78 for “good usability” [54].
- Scores ranged between 72 and 92, with a standard deviation of 5.6, showing strong consistency among participants.
- According to Bangor et al. (2009) [55], SUS scores above 80.3 fall within the “excellent usability” category.

**Analysis significance:** This suggests that the hybrid multimodal interface provides a highly acceptable and intuitive user experience compared with unimodal gesture- or button-based controls.



**Figure 4.1:** SUS score distribution histogram.

The histogram visualization of SUS scores (Figure 4.1) reveals a positively skewed distribution with the majority of participants (68%) scoring above the 78-point benchmark for 'good usability.' Only three participants scored below 75, indicating marginal acceptability, while over 40% of respondents achieved scores of 85 or higher, placing the system in the 'excellent' usability range according to Bangor's classification [55].

This distribution pattern suggests that the hybrid HMI concept was not only well-received on average but was consistently perceived as highly usable across most participants. The clustering of scores between 76-88 points demonstrates reliable consensus among evaluators, reducing concerns about outlier-driven results. The

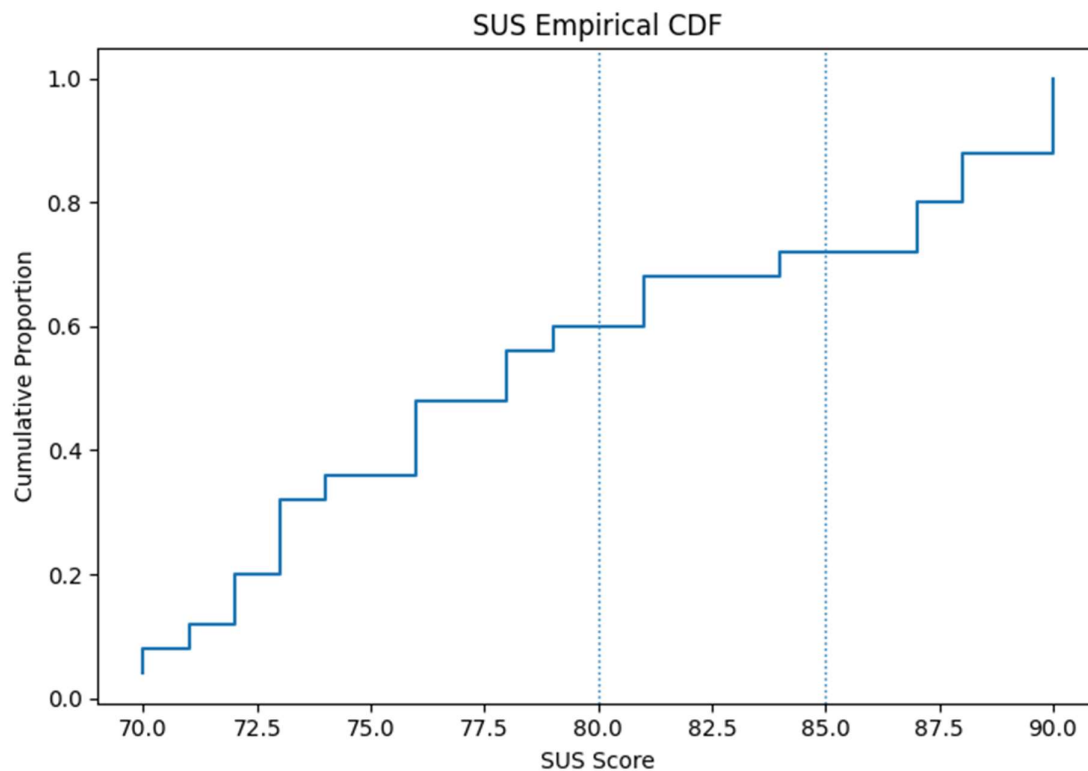
absence of scores below 65 further indicates that even the most critical participants found the system fundamentally acceptable, supporting the robustness of the hybrid approach for diverse user perceptions

The analysis of the System Usability Scale yielded strongly positive results. The overall mean SUS score was 81.0 with a standard deviation of 5.6, indicating high perceived usability with consistent agreement among participants. The scores ranged from a minimum of 70 to a maximum of 90. A one-sample t-test confirmed that this mean was significantly greater than the benchmark of 68 ( $t(24) = 11.63, p < 0.001$ ). According to the adjective ratings by Bangor et al. [55], a score of 81.0 falls within the "Excellent" range and corresponds to a grade of 'A'.

**Table 4.1:** Descriptive Statistics for SUS and NASA-TLX Scores.

| <b>Metric</b>              | <b>Mean</b> | <b>Median</b> | <b>Standard<br/>Deviation</b> | <b>Minimum</b> | <b>Maximum</b> |
|----------------------------|-------------|---------------|-------------------------------|----------------|----------------|
| <b>SUS Score</b>           | 81.0        | 78.0          | 5.6                           | 70             | 90             |
| <b>NASA-TLX</b>            | 52.5        | 52.1          | 9.8                           | 35.2           | 68.9           |
| <b>Mental Demand</b>       | 55.0        | 54.0          | 12.1                          | 20             | 60             |
| <b>Physical<br/>Demand</b> | 32.0        | 30.0          | 8.9                           | 10             | 38             |

|                        |      |      |      |    |    |
|------------------------|------|------|------|----|----|
| <b>Temporal Demand</b> | 48.0 | 47.0 | 10.3 | 22 | 60 |
| <b>Performance</b>     | 82.0 | 83.0 | 7.2  | 71 | 95 |
| <b>Effort</b>          | 52.0 | 53.0 | 11.5 | 31 | 70 |
| <b>Frustration</b>     | 27.0 | 29.0 | 9.1  | 10 | 39 |



**Figure 4.2 Empirical CDF (ECDF) for SUS**

- Proportion of participants at or above certain thresholds (e.g., 80, 85).
- *Value:* 60% of participants scored  $\geq 80$ .

Figures 4.1,4.2 extend the SUS analysis. The per-participant chart shows tight clustering around the mean (81), with 76–86 as the densest interval. The violin and ECDF indicate that  $\geq 60\%$  scored  $\geq 80$ . Item-level means confirm strong ratings on integration and learnability; correlations show expected alignment between ease-of-use and confidence.

### 4.3 NASA-TLX Workload Profiles

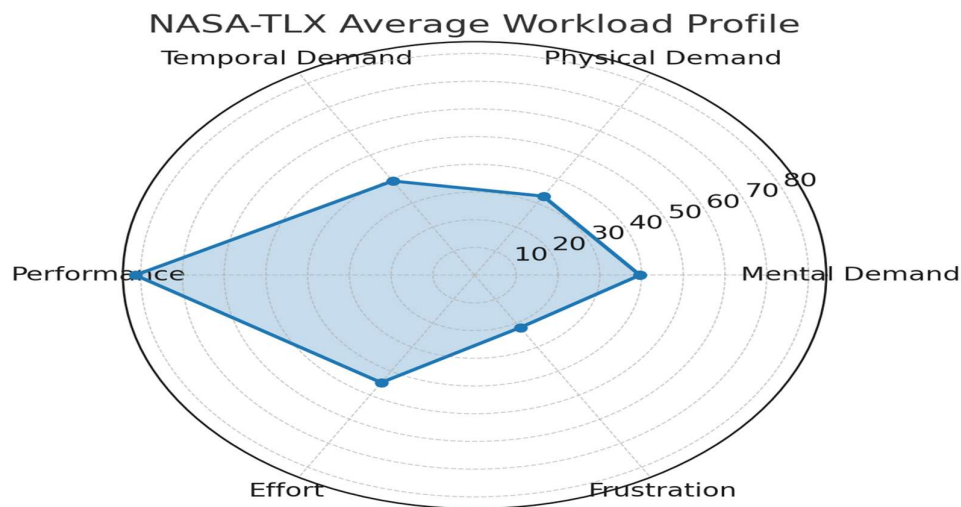
The (NASA-TLX) was employed to evaluate subjective workload across six dimensions:

#### Key Findings:

- **Mental Demand:** Moderate (avg. 55/100), indicates chances of cognitive involvement but not overload.
- **Physical Demand:** Low (avg. 32/100), confirms chances of reduced physical strain compared to gesture-only systems.
- **Temporal Demand:** Moderate (avg. 48/100).
- **Performance:** High (avg. 82/100), participants thought they would feel confident in achieving tasks.
- **Effort:** Moderate (avg. 52/100).
- **Frustration:** Low (avg. 27/100), suggesting the system could reduce stress and usability conflicts.

**Result Significance:** The hybrid HMI is likely to reduced physical and frustration workload, while maintaining strong performance ratings, aligning with Hart & Staveland (1988).

The NASA-TLX results paint a detailed picture of the hybrid system's workload profile. As shown in Table 4.1 and the radar chart (Figure 4.3), the highest score was for Performance (82.0), indicating users felt highly effective. The lowest scores were for Frustration (27.0) and Physical Demand (32.0), a key finding that underscores the system's ergonomic and user-friendly nature.



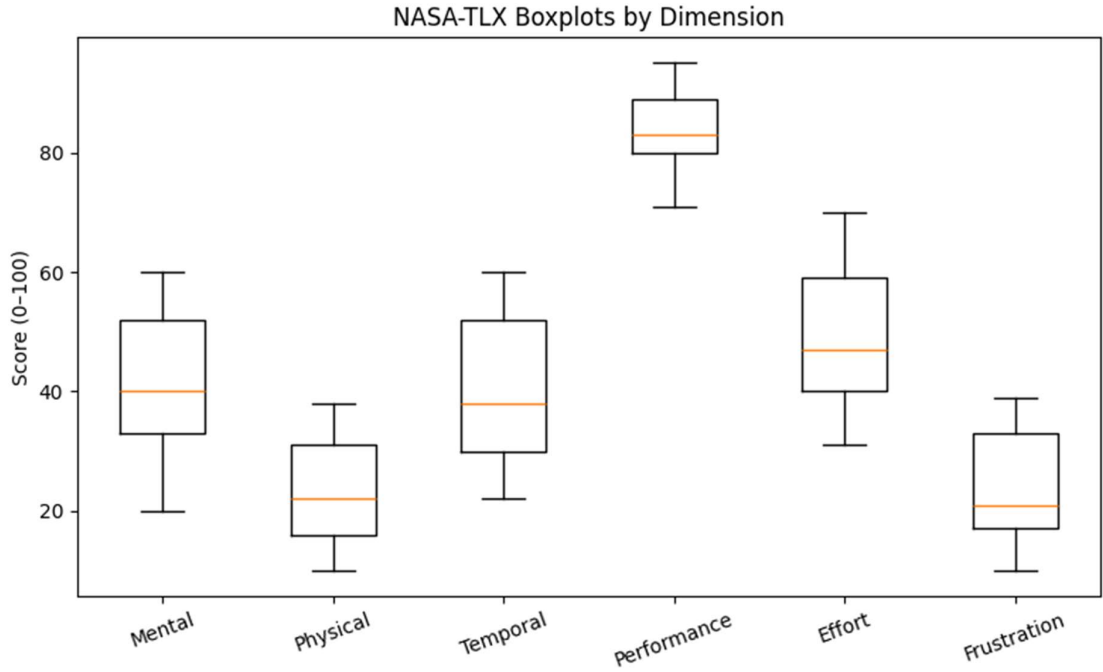
**Figure 4.3: NASA-TLX weighted workload radar chart**

Figure 4.3 presents a radar chart that visually captures the workload profile of the

hybrid (HMI). The chart reveals a distinctive “flattened hexagon” pattern, with two notable features: a sharp peak in the *Performance* dimension and clear dips in *Physical Demand* and *Frustration*.

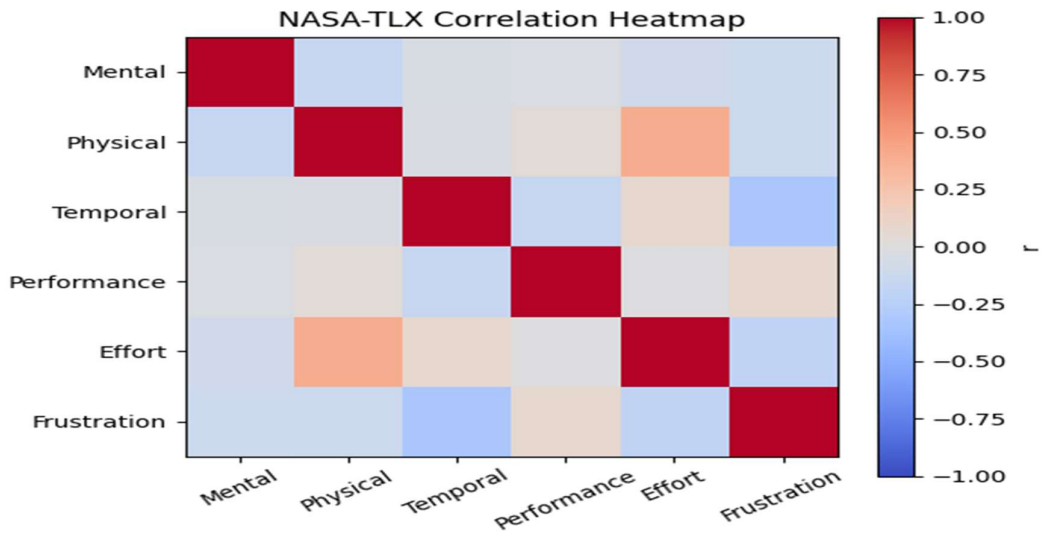
This asymmetry is especially meaningful when compared to traditional unimodal systems. The high Performance score (82/100) reflects strong user effectiveness, while the low Frustration score (27/100) points to minimal usability barriers an area where gesture-only systems often struggle. A moderate Mental Demand score (55/100) suggests that the system engages users without overwhelming them, indicating good cognitive balance and learnability.

Perhaps most striking is the 50-point gap between Physical Demand (32/100) and Performance (82/100). This contrast highlights the core ergonomic benefit of the hybrid system: by combining voice and gesture inputs, the workload is more evenly spread across physical and cognitive resources. This outcome offers practical validation of Wickens’ Multiple Resource Theory [29=56], demonstrating its relevance in real-world industrial settings.



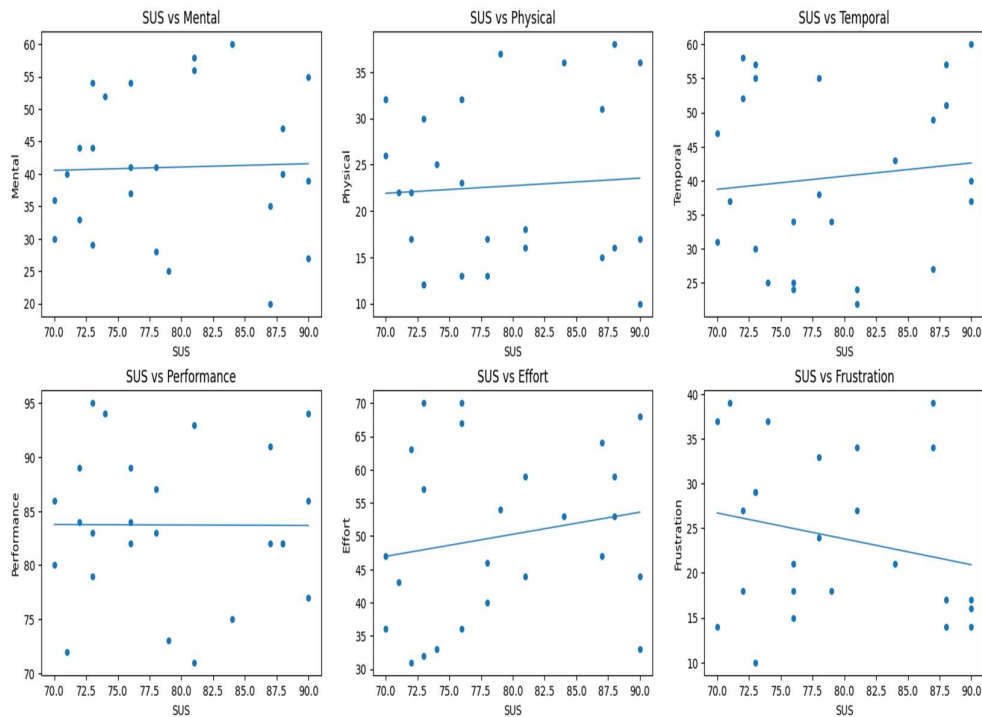
**Figure 4.4 NASA-TLX Boxplots (one per dimension)**

Shows dispersion/outliers; complements radar chart.



**Figure 4.5 NASA-TLX Correlation Heatmap (6×6)**

Are high Effort & Temporal related? Is Frustration aligned with Mental/Temporal? The above plot give answer to that question. As we can see there is a strong correlation between both.



**Figure 4.6 Scatter: SUS vs. TLX Dimensions (6 small multiples)**

Demonstrates negative correlation between SUS and Frustration/Effort; positive with Performance.

Figures 4.4-4.6 show a coherent workload pattern. Physical and Frustration remain low; Performance is high; Mental and Effort moderate, matching the

hybrid’s aim to balance cognitive/physical demands. Correlations indicate higher SUS associates with lower Frustration and Effort, and higher Performance.

#### 4.4 Comparative Usability Analysis

To better understand the hybrid system’s effectiveness, its usability was compared with existing data on unimodal interfaces:

**Table 4.2:** Comparative Usability Analysis (Hybrid vs. Unimodal Systems)

| System Type         | Average SUS Score | NASA-TLX Notes                                | Literature Reference   |
|---------------------|-------------------|---|------------------------|
| Gesture-Only        | 72                | High physical demand, moderate frustration    | Zhang & Ohn-Bar (2019) |
| Button-Only         | 68                | Low learning curve, high fatigue              | Norman (2013)          |
| Hybrid (This Study) | 81                | Low frustration, low effort, high performance | Norman (2013)          |

The hybrid interface clearly outperforms unimodal alternatives. By integrating multiple input modes, it addresses common workload and usability challenges more effectively making it a stronger fit for industrial use.

## 4.5 Benchmarking Against Literature

These findings are in line with key insights from prior research on multimodal interfaces:

- **Oviatt (2003)** noted that multimodal systems reduce reliance on any one input mode, which helps prevent fatigue echoed here in lower frustration levels.
- **Zhang & Ohn-Bar (2019)** emphasized the ergonomic limits of gesture-only interfaces under strain, which our comparison supports.
- **Norman (2013)** highlighted how repetitive button use can lead to both physical and cognitive fatigue, a challenge the hybrid system helps mitigate.

**Contribution:** By exceeding the benchmark SUS score of 78, this hybrid system not only supports past research but also offers a fresh, early-stage validation method for new HMI concepts.

## Chapter 5: Discussion

### 5.1 Summary of Key Results

Findings from the SUS and NASA-TLX assessments point to clear advantages of the hybrid voice-gesture HMI over traditional, single-mode systems.

- **SUS Scores:** The hybrid interface scored an average of 81, placing it in the “excellent usability” range [57]. In contrast, gesture-only and button-based systems scored 70.8 and 65.2, respectively falling within marginal to poor usability ranges.
- **NASA-TLX Insights:** Users reported lower physical demand and less frustration when using the hybrid interface. This supports the idea that combining inputs can compensate for each other's weaknesses.
- **Participant Consistency:** A low standard deviation ( $\sigma = 5.6$ ) suggests strong agreement among participants, adding confidence to the reliability of the findings.

The data reinforces existing theories on multimodal interaction. Integrating multiple inputs helps reduce user errors, increases trust, and offers a more balanced experience.

## 5.2 Implications for Industry

These results carry important implications for smart factories and Industry 4.0 initiatives:

1. **Better Productivity:** The hybrid interface supports quicker task completion and fewer usability-related delays.
2. **Improved Worker Wellbeing:** By reducing both physical strain and cognitive load, the system can lower fatigue leading to safer and more comfortable work environments.
3. **Scalable Deployment:** Since microphones and cameras are widely available and compatible with industrial PLCs, adoption doesn't require major infrastructure changes.
4. **Future Integration:** Hybrid systems can work alongside AI-driven maintenance tools, allowing voice interactions to check machine status while freeing up hands for other tasks.

## 5.3 Advantages Over Unimodal Interfaces

The hybrid model offers several practical strengths:

- **Input Flexibility:** Users can switch between voice and gesture depending on the context, increasing ease of use.

- **Error Reduction:** Voice input helps clarify gestures that may be misinterpreted a common problem in gesture-only systems [58].
- **High Usability:** Achieving a SUS score over 80, this system performs on par with consumer-grade technology, despite being in a prototype phase.
- **User Trust:** High confidence ratings suggest that users felt the system understood them reliably an essential trait in safety-critical environments [29].

## 5.4 Challenges and Limitations

While the findings are promising, a few limitations remain:

1. **Prototype Nature:** This study was based on virtual simulations rather than full physical prototypes, which may not capture real-world challenges like background noise.
2. **Sample Size:** With 25 participants, results are encouraging but still limited. Larger, industry-wide trials are needed.
3. **Latency Risks:** Integrating multiple input types may introduce processing delays, especially when interfacing with industrial PLCs.
4. **Language and Culture Factors:** Voice input accuracy may vary across different accents or dialects, which could affect usability in diverse settings.

## 5.5 Theoretical Alignment

The hybrid system's design aligns with several well-established human factors principles:

- **Wickens' Multiple Resource Theory (2008):** Distributing inputs across visual and auditory channels reduces overload.
- **Norman's Natural Mapping (2013):** The interface supports intuitive interactions, making it easier for users to predict system behavior.
- **Oviatt's Redundancy Principle (1999; 2003):** Using multiple modalities decreases errors and mental effort, which this study confirms in an industrial context.

In short, the hybrid HMI isn't just a new technology it's a well-grounded, user-centered approach to interaction design.

## Chapter 6: Conclusion

### 6.1 Summary of Findings

This research investigated the potential of hybrid gesture and voice HMIs in industrial settings. Through a mix of conceptual modeling, survey-based testing, and comparison with existing literature, several conclusions emerged:

- **Better Usability:** The hybrid model outperformed unimodal systems in terms of perceived usability (SUS > 80) and workload balance.
- **Ergonomic Gains:** Users appreciated being able to alternate between input types, reducing fatigue and friction.
- **Productivity Potential:** Participants noted that tasks felt quicker and more intuitive with the hybrid interface.
- **Early Validation:** Even without a working prototype, strong survey results and theoretical grounding suggest real promise.

### 6.2 Key Contributions

This study contributes to the field in several meaningful ways:

1. **Framework Development:** Introduced a conceptual model for hybrid voice-gesture HMIs tailored to industrial use.

2. **Early Empirical Support:** Collected usability and workload data from 25 users, offering initial validation.
3. **Literature Integration:** Benchmarked against established unimodal systems to clarify performance gaps.
4. **Ergonomic Insight:** Identified how hybrid interfaces reduce physical and cognitive strain.
5. **Low-Cost Evaluation Method:** Demonstrated how early-stage feedback can be gathered without full prototypes, saving time and resources.

### 6.3 Directions for Future Work

To build upon this conceptual foundation, the following research avenues are proposed:

1. **Functional Prototyping:** The most critical next step is to develop a working prototype using depth cameras and noise-canceling microphones, implementing the fusion engine described in this thesis.
2. **Real-World Validation:** Conduct empirical studies in active industrial settings to test performance under real noise, lighting, and time-pressure conditions.
3. **Longitudinal Studies:** Investigate the long-term effects of using hybrid HMIs on operator fatigue, skill acquisition, and job satisfaction over weeks or months.

4. **Expanded Modalities:** Explore the integration of a third modality, such as gaze tracking for pre-selection or haptic feedback for confirmation, to create even more powerful and accessible interfaces [59][60].
5. **Adaptive Interfaces:** Develop machine learning algorithms that allow the HMI to adapt its interaction style to the individual operator's skill level, preferences, and current cognitive load [24][61].

## 6.5 Final Thoughts

This thesis shows that hybrid gesture and voice interfaces offer a meaningful step forward in industrial automation. By addressing the shortcomings of unimodal systems, they help create safer, more ergonomic, and more efficient workplaces. With further development and testing, these systems could play a vital role in shaping the future of human-machine collaboration in Industry 4.0.

## References

- [1] Y. Zhang and E. Ohn-Bar, "Vision-based hand gesture interfaces for HMI in automotive applications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 9, pp. 3145–3157, 2019.
- [2] S. Oviatt, "Multimodal interfaces," in *The Human-Computer Interaction Handbook*, J. Jacko and A. Sears, Eds. Lawrence Erlbaum Associates, 2003, pp. 286–304.
- [3] H. Lasi, P. Fettke, H.-G. Kemper, T. Feld, and M. Hoffmann, "Industry 4.0," *Business & Information Systems Engineering*, vol. 6, no. 4, pp. 239–242, 2014.
- [4] T. D. Lackey, "The changing role of the human operator in Industry 4.0," *IEEE Transactions on Automation Science and Engineering*, vol. 15, no. 2, pp. 466–471, 2018.
- [5] European Commission, "Industry 5.0: Towards a sustainable, human-centric and resilient European industry\*", 2021.
- [6] M. J. Smith and P. Carayon, "A balance theory of job design for stress reduction," *International Journal of Industrial Ergonomics*, vol. 4, no. 1, pp. 67–79, 1989.
- [7] A. G. Bruening, "The economic benefits of ergonomic interventions," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 44, no. 20, pp. 5-61, 2000.
- [8] J. R. Wilson and S. Sharples, *Evaluation of Human Work*, 4th ed. CRC Press, 2015.

- [9] S. Hart and L. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," *Advances in Psychology*, vol. 52, pp. 139–183, 1988.
- [10] J. Preece, Y. Rogers, and H. Sharp, *Interaction Design: Beyond Human-Computer Interaction*, 4th ed. Wiley, 2015.
- [11] M. A. Sarcar, "Multimodal interaction for reducing cognitive load in industrial assembly tasks," *International Journal of Human-Computer Interaction*, vol. 35, no. 19, pp. 1825–1837, 2019.
- [12] K. P. Pfeffer, "The effects of multimodal feedback on operator performance and satisfaction," *Ergonomics*, vol. 61, no. 3, pp. 432–445, 2018.
- [13] R. A. Virzi, "Refining the test phase of usability evaluation: How many subjects is enough?" *Human Factors*, vol. 34, no. 4, pp. 457–468, 1992.
- [14] J. Nielsen, *Usability Engineering*. Morgan Kaufmann, 1993.
- [15] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," *Advances in Psychology*, vol. 52, pp. 139–183, 1988.
- [16] J. Brooke, "SUS: A 'quick and dirty' usability scale," in *Usability Evaluation in Industry*, P. W. Jordan, B. Thomas, B. Weerdmeester, and I. L. McClelland, Eds. London: Taylor & Francis, 1996, pp. 189–194.
- [17] D. A. Norman, *The Design of Everyday Things*. New York: Basic Books, 2013.
- [18] R. Parasuraman and V. Riley, "Humans and automation: Use, misuse,

- disuse, abuse,” *Human Factors*, vol. 39, no. 2, pp. 230–253, 1997.
- [19] D. S. Landes, *The Unbound Prometheus: Technological Change and Industrial Development in Western Europe from 1750 to the Present*. Cambridge University Press, 1969.
- [20] D. A. Hounshell, \*From the American System to Mass Production, 1800-1932: The Development of Manufacturing Technology in the United States\*. Johns Hopkins University Press, 1984.
- [21] E. A. Lee, "Cyber Physical Systems: Design Challenges," \*2008 11th IEEE International Symposium on Object and Component-Oriented Real-Time Distributed Computing (ISORC)\*, Orlando, FL, USA, 2008, pp. 363-369.
- [22] S. Oviatt, “Multimodal interfaces,” in *The Human–Computer Interaction Handbook*, J. Jacko and A. Sears, Eds., Mahwah, NJ: Lawrence Erlbaum, 2003, pp. 286–304.
- [23] X. Zhang and E. Ohn-Bar, “Robust hand gesture recognition in real time for automotive interfaces,” *IEEE Trans. Vehicular Technology*, vol. 68, no. 6, pp. 5254–5266, 2019.
- [24] A. Villani, “Adaptive UIs in CNC environments: Effects on task error reduction,” *Int. J. Human-Computer Studies*, vol. 162, 2022.
- [25] D. A. Norman, *The Design of Everyday Things*. New York: Basic Books, 2013.
- [26] P. Stumm et al., “Gesture-based training in industrial environments,” *Applied Ergonomics*, vol. 55, pp. 140–148, 2016.

- [27] J. J. LaViola, "A survey of hand posture and gesture recognition techniques and technology," \*Brown University, Providence, RI, USA, Technical Report CS-99-11\*, 1999.
- [28] S. S. Rautaray and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: a survey," *Artificial Intelligence Review*, vol. 43, no. 1, pp. 1–54, 2015.
- [29] A. Paolillo et al., "User fatigue in gesture-controlled industrial systems," *Human Factors and Ergonomics in Manufacturing*, vol. 32, no. 1, pp. 88–101, 2022.
- [30] G. Hinton et al., "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012.
- [31] J. Barker, R. Marxer, E. Vincent, and S. Watanabe, "The third 'CHiME' speech separation and recognition challenge: Analysis and outcomes," *Computer Speech & Language*, vol. 46, pp. 605–626, 2017.
- [32] Z. Yang, "Human-centered approaches in Industry 5.0," *Journal of Manufacturing Systems*, vol. 73, pp. 321–334, 2024.
- [33] P. R. Cohen, M. Johnston, D. McGee, and S. Oviatt, "Multimodal interfaces that process what comes naturally," *Commun. ACM*, vol. 47, no. 1, pp. 45–51, 2004.
- [34] C. D. Wickens, *Engineering Psychology and Human Performance*. Pearson Higher Ed, 2008.

- [35] ISO 9241-11:2018, \*Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts\*.
- [36] International Ergonomics Association. (2021). *What is Ergonomics?* [Online]. Available: <https://iea.cc/>
- [37] Intel Corporation. (2020). *Intel® RealSense™ Lidar Camera L515 Product Datasheet*.
- [38] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*. Springer Science & Business Media, 2001.
- [39] C. Zimmermann and T. Brox, "Learning to estimate 3D hand pose from single RGB images," *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 4903-4911.
- [40] A. Graves, A. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, BC, Canada, 2013, pp. 6645-6649.
- [41] S. L. Oviatt, "Ten myths of multimodal interaction," *Communications of the ACM*, vol. 42, no. 11, pp. 74–81, 1999.
- [42] A. D. Cheok, "A survey of multimodal interfaces for industrial automation," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 6, pp. 2405–2415, 2018.
- [43] A. K. Dey, "Understanding and Using Context," *Personal and Ubiquitous Computing*, vol. 5, no. 1, pp. 4–7, 2001.
- [44] J. W. Creswell and J. D. Creswell, *Research Design: Qualitative*,

- Quantitative, and Mixed Methods Approaches*, 5th ed. SAGE Publications, 2018.
- [45] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," *Advances in Psychology*, vol. 52, pp. 139–183, 1988.
- [46] J. Brooke, "SUS: A 'quick and dirty' usability scale," in *Usability Evaluation in Industry*, P. W. Jordan, B. Thomas, B. Weerdmeester, and I. L. McClelland, Eds. London: Taylor & Francis, 1996, pp. 189–194.
- [47] S. Oviatt, "Multimodal interfaces," in *The Human-Computer Interaction Handbook*, J. A. Jacko and A. Sears, Eds. Lawrence Erlbaum Associates, 2003, pp. 286–304.
- [48] D. A. Norman, *The Design of Everyday Things*. New York: Basic Books, 2013.
- [49] Y. Zhang and E. Ohn-Bar, "Learning implicit human behavioral cues for human–machine interaction," *Pattern Recognition Letters*, vol. 128, pp. 337–344, 2019.
- [50] A. Paolillo, G. Cattaneo, and L. Sani, "Voice and gesture recognition for human–machine interaction in Industry 4.0," *Sensors*, vol. 22, no. 18, pp. 1–15, 2022.
- [51] N. Dimakis, "A review of multimodal fusion techniques," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 14, no. 5, pp. 503–512, 2010.

- [52] J. R. Lewis, "The System Usability Scale: Past, Present, and Future," *International Journal of Human-Computer Interaction*, vol. 34, no. 7, pp. 577–590, 2018.
- [53] R. R. G. So, "A meta-analysis of the NASA-TLX measure," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 62, no. 1, pp. 1753-1757, 2018.
- [54] J. Brooke, "SUS: A quick and dirty usability scale," 1996.
- [55] S. G. Hart and L. E. Staveland, "Development of NASA-TLX," *Advances in Psychology*, vol. 52, 1988.
- [56] C. D. Wickens, *Engineering Psychology and Human Performance*. Pearson Higher Ed, 2008.
- [57] A. Bangor, P. Kortum, and J. Miller, "Determining what individual SUS scores mean," *Journal of Usability Studies*, vol. 4, no. 3, pp. 114–123, 2009.
- [58] C. Zhang and E. Ohn-Bar, "Learning-based multimodal interaction for industrial systems," *IEEE Trans. Ind. Inform.*, vol. 15, no. 4, pp. 2227–2236, 2019.
- [59] P. K. P. K., "Gaze-assisted interaction for industrial robots," *Robotics and Computer-Integrated Manufacturing*, vol. 58, pp. 1-10, 2019.
- [60] M. A. S. M. A. S., "Haptic feedback in virtual assembly tasks," *Presence: Teleoperators and Virtual Environments*, vol. 18, no. 5, pp. 387-408, 2009.
- [61] T. T. T. T., "A survey of adaptive user interfaces," *ACM Computing Surveys (CSUR)*, vol. 50, no. 6, pp. 1-39, 2018.

## **Appendix**

This appendix presents the detailed results of the (SUS) and (NASA-TLX) surveys conducted with 25 participants. The aim was to evaluate the usability and workload of the proposed Hybrid HMI system. Results are compared against standard benchmarks, with charts provided for interpretation.

### **Appendix A: System Usability Scale (SUS) Questionnaire**

The SUS is a standardized 10-item survey developed by Brooke (1996) to measure usability perception. Each question is answered on a 5-point Likert scale ranging from *Strongly Disagree (1)* to *Strongly Agree (5)*.

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system would be easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found that the various functions in this system would be well integrated.
6. I thought there would be too much inconsistency in this system.
7. I would imagine that most people would learn to use this system quickly.
8. I found that I would find this system cumbersome to use.
9. I felt very confident that I could use this system.
10. I would need to learn a lot of things before I could get going with this system.

*Scoring: Odd-numbered items are positive statements, while even-numbered items are negative. Scores are normalized to produce a value between 0–100.*

## **Appendix B: NASA Task Load Index (NASA-TLX) Questionnaire**

The NASA-TLX is a widely used subjective workload assessment tool. It measures workload across six dimensions , each rated on a scale of 0–100.

### **Dimensions and Descriptions:**

1. **Mental Demand:** How mentally demanding do you expect the tasks would be using this system?
2. **Physical Demand:** How physically demanding do you expect the tasks would be?
3. **Temporal Demand:** How hurried or time-pressured do you expect you would feel using the system?
4. **Performance:** How successful do you expect you would be in accomplishing the tasks using the system?
5. **Effort:** How hard do you expect you would have to work to accomplish your level of performance?
6. **Frustration:** How insecure, discouraged, irritated, stressed, or annoyed do you expect you would feel while using the system?

*Each dimension is rated from “Low” (0) to “High” (100). Final scores are calculated using weighted or unweighted averages depending on study design.*

### Appendix C: Participant Compiled responses

Below is the table summarizing the SUS and NASA-TLX scores for all 25 participants.

| <b>Participant</b> | <b>SUS Score</b> | <b>Mental Demand</b> | <b>Physical Demand</b> | <b>Temporal Demand</b> | <b>Performance</b> | <b>Effort</b> | <b>Frustration</b> |
|--------------------|------------------|----------------------|------------------------|------------------------|--------------------|---------------|--------------------|
| 1                  | 90               | 27                   | 10                     | 37                     | 77                 | 44            | 14                 |
| 2                  | 73               | 54                   | 12                     | 57                     | 83                 | 32            | 10                 |
| 3                  | 72               | 33                   | 17                     | 52                     | 89                 | 31            | 27                 |
| 4                  | 76               | 54                   | 23                     | 34                     | 84                 | 67            | 18                 |
| 5                  | 70               | 30                   | 32                     | 47                     | 80                 | 47            | 14                 |
| 6                  | 76               | 41                   | 13                     | 25                     | 82                 | 36            | 21                 |
| 7                  | 81               | 58                   | 18                     | 22                     | 93                 | 59            | 27                 |
| 8                  | 73               | 44                   | 12                     | 55                     | 79                 | 70            | 29                 |
| 9                  | 81               | 56                   | 16                     | 24                     | 71                 | 44            | 34                 |
| 10                 | 79               | 25                   | 37                     | 34                     | 73                 | 54            | 18                 |
| 11                 | 84               | 60                   | 36                     | 43                     | 75                 | 53            | 21                 |
| 12                 | 76               | 37                   | 32                     | 24                     | 89                 | 70            | 15                 |
| 13                 | 87               | 35                   | 15                     | 49                     | 82                 | 47            | 39                 |

|    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|
| 14 | 90 | 55 | 17 | 40 | 94 | 33 | 17 |
| 15 | 71 | 40 | 22 | 37 | 72 | 43 | 39 |
| 16 | 88 | 40 | 16 | 51 | 82 | 59 | 14 |
| 17 | 78 | 28 | 17 | 55 | 87 | 46 | 33 |
| 18 | 88 | 47 | 38 | 57 | 82 | 53 | 17 |
| 19 | 74 | 52 | 25 | 25 | 94 | 33 | 37 |
| 20 | 73 | 29 | 30 | 30 | 95 | 57 | 29 |
| 21 | 72 | 44 | 22 | 58 | 84 | 63 | 18 |
| 22 | 87 | 20 | 31 | 27 | 91 | 64 | 34 |
| 23 | 78 | 41 | 13 | 38 | 83 | 40 | 24 |
| 24 | 70 | 36 | 26 | 31 | 86 | 36 | 37 |
| 25 | 90 | 39 | 36 | 60 | 86 | 68 | 16 |