Cross-Modal Mapping and Wearable Multimodal Data in Human-Computer Interaction: A Survey

Abstract

This survey examines the transformative potential of cross-modal mapping and wearable multimodal data in enhancing Human-Computer Interaction (HCI). By integrating diverse sensory modalities such as motion, text, and physiological signals, these technologies enable more intuitive and adaptive interaction systems. The survey highlights key advancements in gesture recognition, emotion detection, and smart healthcare applications, showcasing the benefits of multimodal integration in improving user experience and system responsiveness. Despite significant progress, challenges remain in data accuracy, device usability, and ethical concerns, necessitating further research into robust models and secure data handling practices. Future directions include the development of user-friendly healthcare technologies, improved gesture recognition frameworks, and enhanced emotion recognition models. By addressing these challenges and leveraging interdisciplinary approaches, the field of HCI can advance towards more effective and user-centered interaction systems, ultimately enhancing the quality of human-computer interactions across various domains.

1 Introduction

1.1 Overview of Cross-Modal Mapping and Wearable Multimodal Data

Cross-modal mapping and wearable multimodal data significantly enhance Human-Computer Interaction (HCI) by integrating diverse data types, including motion, text, and physiological signals. This integration improves user experience by fostering intuitive and context-aware interactions while addressing limitations inherent in traditional methods within dynamic environments [1]. Vision-based Multimodal Interfaces (VMIs), combined with multimodal AI technologies, notably enhance context awareness, resulting in adaptive and responsive systems [2].

User-centered design is pivotal in intelligent HCI (iHCI), necessitating a paradigm shift in research and design to better meet user needs through AI integration [3]. This is especially relevant in smart healthcare devices, where HCI principles improve usability and engagement [4]. The ARISES app exemplifies this approach by facilitating Type 1 diabetes self-management through real-time data integration from wearables [5].

Wearable devices are crucial for continuous monitoring of human activity and biological signals, addressing health challenges linked to sedentary lifestyles and aging [6]. A systematic categorization of deep learning methods for wearable-based Human Activity Recognition (HAR) illustrates the potential of these technologies to enhance understanding of activity patterns and user experience [7]. Moreover, algorithms that detect early COVID-19 signs using wearable data highlight the significance of these technologies in early medical intervention [8].

Automatic emotion recognition in HCI remains a critical area, with existing methods often relying on simplistic data representations that may lead to overfitting [9]. The integration of multimodal data—visual, thermal, and audio—is increasingly important in affective computing, prompting novel

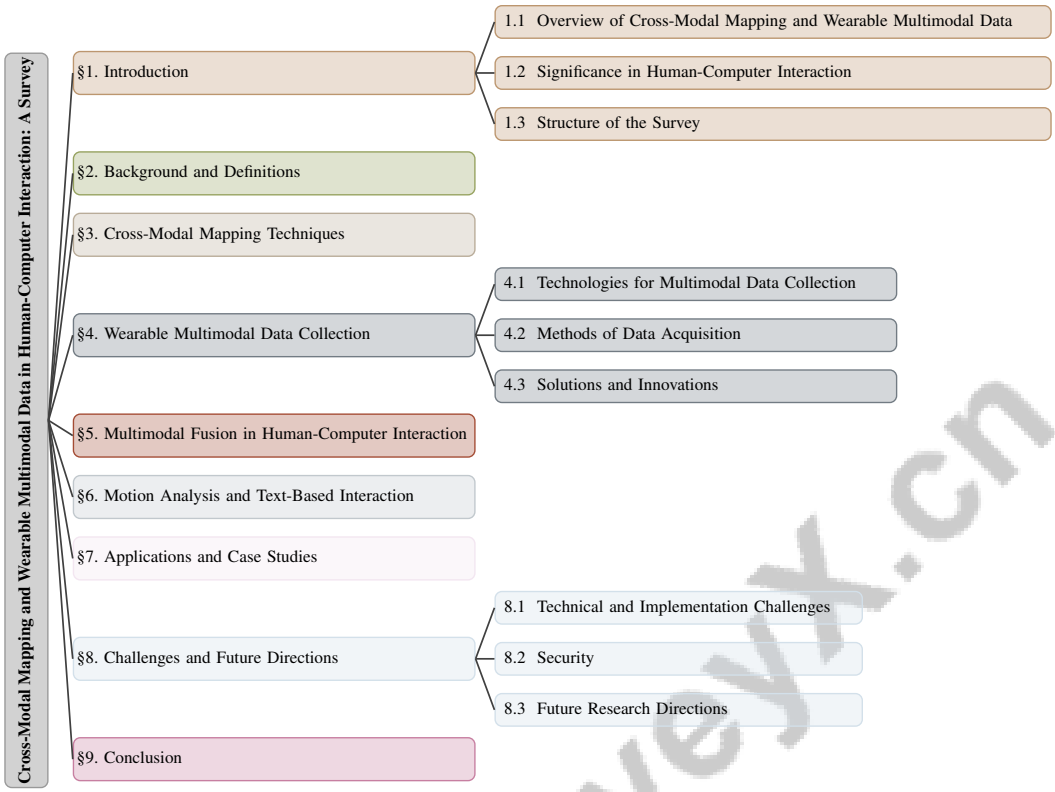


Figure 1: chapter structure

approaches for emotion recognition. The evolution of sensory earables from basic audio devices to advanced health monitoring platforms underscores the ongoing need for enhanced performance and real-time processing capabilities in wearable technology [10].

Cross-modal mapping and wearable multimodal data are essential for advancing HCI, enabling natural and effective interactions. These technologies not only enrich user experience but also pave the way for innovative applications across fields such as gesture recognition and the development of believable movements for conversational agents [11]. The application of natural language processing (NLP) methods to model interactive behavior further enhances task recognition in HCI, showcasing the interdisciplinary integration of these technologies [12].

1.2 Significance in Human-Computer Interaction

The integration of cross-modal mapping and wearable multimodal data significantly enhances HCI by fostering intuitive, adaptive, and personalized user experiences. These technologies transcend the limitations of traditional unimodal interfaces by incorporating diverse data types, such as graphical, voice-based, and immersive interfaces, which are vital for enriching user interactions [13]. Natural User Interfaces (NUIs) are particularly instrumental in providing intuitive interaction methods, thus improving user experiences in mobile applications [14].

The application of visual and haptic feedback through pseudo-haptics can greatly enhance user experiences in virtual environments without relying on physical haptic devices [15]. This method addresses interaction challenges and highlights the potential for innovative interface designs that leverage multimodal feedback. Additionally, Electromyography (EMG), particularly facial EMG, plays a crucial role in everyday HCI scenarios by indexing emotional valence, enhancing interaction dynamics [16].

Incorporating user needs into Augmentative and Alternative Communication (AAC) technologies is essential, as poor design can lead to tool abandonment [17]. Adhering to user-centered design principles enables HCI technologies to better accommodate diverse user needs, enhancing accessibility

and usability. The proposed VAD model, which bridges discrete and continuous emotion detection, exemplifies how nuanced emotion vocabularies can improve affective computing applications [18].

Exploring the influence of science fiction on HCI research underscores its role in shaping future technologies and interactions, emphasizing the importance of interdisciplinary exploration in advancing HCI innovations [19]. The SimplyMime framework integrates diverse data types for enhanced human-computer interaction, demonstrating the significance of multimodal integration in simplifying user experiences [20].

The significance of these technologies in HCI lies in their ability to create more natural, efficient, and personalized interactions. The integration of diverse modalities not only enhances user experience but also facilitates innovative applications across various domains. For instance, the ARISES app addresses high abandonment rates of diabetes health applications through a user-friendly interface, underscoring the importance of user-centered design in healthcare [5]. Similarly, VMIs enhance context awareness by interpreting user intentions and environmental information [2]. The ongoing evolution of these systems highlights the necessity for interdisciplinary research and development to overcome existing limitations and enhance HCI capabilities.

1.3 Structure of the Survey

This survey is organized into several key sections, each addressing distinct aspects of cross-modal mapping and wearable multimodal data in HCI. The introductory section provides a comprehensive overview of the significance of these technologies in enhancing HCI, setting the stage for a deeper exploration of their roles and applications. The background section delves into core concepts and definitions, establishing a foundational understanding of the interdisciplinary integration of diverse data types and their relevance to HCI.

Subsequent sections focus on specific technical aspects, including cross-modal mapping techniques, where gesture recognition and synthesis techniques are examined, alongside innovative models and algorithms that facilitate these processes. The discussion on wearable multimodal data collection explores the technologies and methods employed in acquiring data from wearable devices, highlighting solutions and innovations that address the challenges in this domain.

The survey further investigates the process of multimodal fusion, emphasizing its importance in enhancing HCI through various techniques and strategies, including the role of hybrid interfaces. The analysis of motion and text-based interactions within wearable multimodal data focuses on three key areas: gesture recognition, encompassing advancements in hand gesture recognition systems utilizing various data modalities; facial and body motion analysis, addressing challenges in emotion recognition and expression subjectivity in video content; and the integration of text-based interactions with other modalities, underscoring the importance of multimodal approaches for enhancing human-computer interaction and educational practices [21, 22, 23].

The applications and case studies section provides real-world examples of successful implementations, covering diverse fields such as healthcare, emotion recognition, augmented and virtual reality, and assistive technologies. The survey concludes with a discussion on the challenges and future directions in the field, identifying technical and implementation challenges, security, privacy, and ethical concerns, as well as emerging trends and research opportunities.The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Core Concepts and Definitions

In Human-Computer Interaction (HCI), the integration of cross-modal mapping and wearable multimodal data is pivotal for fostering intuitive and context-aware user experiences. Cross-modal mapping synthesizes auditory, visual, and textual data into comprehensive representations, facilitating tasks like matching, verification, and retrieval, thus overcoming the constraints of unimodal systems [5]. These processes are theoretically underpinned by collaborative cognitive systems, situation awareness, and intelligent agent theories, guiding the implementation of complex interactions [3].

The advent of Omni multimodal representation models marks significant progress by unifying multiple pre-trained models to handle diverse inputs, addressing the intricacies of multimodal data [4]. This

approach is particularly pertinent in smart healthcare, where a four-layer architecture—comprising sensing, communication, data integration, and application layers—supports seamless user-centered design.

Wearable technology is central to multimodal data acquisition, focusing on continuous monitoring of aerobic activities via fitness trackers. However, performance variability among brands necessitates standardization and improvement [24]. The evolution of sensory earables, traditionally reliant on external processing, highlights the need to mitigate inefficiencies, latency, and privacy concerns to enhance real-time processing.

Human Activity Recognition (HAR) is integral to wearable multimodal data, identifying Activities of Daily Living (ADLs) from sensor inputs. Deep learning offers opportunities and challenges in improving HAR's accuracy, demanding robust models capable of handling diverse data [4].

Gesture recognition is vital for intuitive interactions, especially in augmented and virtual reality, where precise micro-gesture recognition is crucial for user engagement. The challenge of real-time hand motion gesture recognition using non-contact capacitive sensing, often affected by noise, underscores the need for advanced sensing technologies [24].

Text-based interactions and computer-mediated communication (CMC) are significant, with tools like Chat-Bot-Kit enhancing chat interaction evaluation. Natural Language Processing (NLP) methods are essential for modeling interactive behavior, including sequential and hierarchical actions akin to natural language structures [4].

Eye-tracking systems are crucial for natural interaction, addressing the limitations of current methods. Moreover, the challenge of emotion recognition from speech and text necessitates high-level feature extraction methods to enhance accuracy [4].

Addressing sedentary behavior in office environments through augmented reality head-mounted displays exemplifies the application of these concepts to promote health. Challenges in immersive environments, such as performer and prop tracking, illustrate the need for innovative solutions to overcome occlusion and tracking issues [24]. These core concepts lay the groundwork for the interdisciplinary integration of diverse data types in HCI, facilitating further exploration and application across various domains.

2.2 Interdisciplinary Integration

The interdisciplinary integration of cross-modal mapping and wearable multimodal data within HCI is essential for developing user-centric, technologically advanced systems. This integration draws from computer science, psychology, linguistics, and cognitive science to enhance usability and functionality. The complexity of human emotions, not fully captured by a single modality, necessitates a multimodal approach, as evidenced by advanced emotion recognition methods utilizing knowledge distillation [3].

In smart healthcare, interdisciplinary collaboration is vital for enhancing usability and effectiveness, requiring healthcare professionals, engineers, and designers to create functional, intuitive systems [4]. HAR exemplifies this approach by integrating machine learning, sensor technology, and HCI principles to improve activity recognition accuracy. The taxonomy categorizing HAR research highlights the diverse sensor types, application domains, and deep learning methodologies employed, underscoring multifaceted challenges and solutions [3].

The integration of multimodal systems in biometric authentication and recognition demonstrates the convergence of technology and user-centric design, ensuring systems are secure, accessible, and efficient [4]. The interdisciplinary nature of user studies bridges computer graphics and HCI, facilitating comprehensive evaluations and effective system development.

This interdisciplinary exploration extends to wearable devices, where challenges such as limited onboard computation hinder real-time data processing. Addressing these limitations demands collaboration across fields to develop efficient solutions. The categorization of research into domains like operating systems and application development platforms further emphasizes the tailored solutions required for wearable devices [3].

In recent years, the exploration of cross-modal mapping techniques has gained significant traction, particularly in the realm of gesture recognition and synthesis. These advancements are pivotal for

enhancing Human-Computer Interaction (HCI), as they facilitate more intuitive and adaptive user experiences. To illustrate this complex landscape, Figure 2 presents a comprehensive overview of the hierarchical structure of these techniques. This figure categorizes key advancements, showcasing innovative models and algorithms alongside their practical applications. Notably, the diagram highlights the integration of vision-based interfaces and memory-augmented architectures, as well as interdisciplinary innovations, thereby emphasizing their critical roles in the development of user-centered interaction systems.

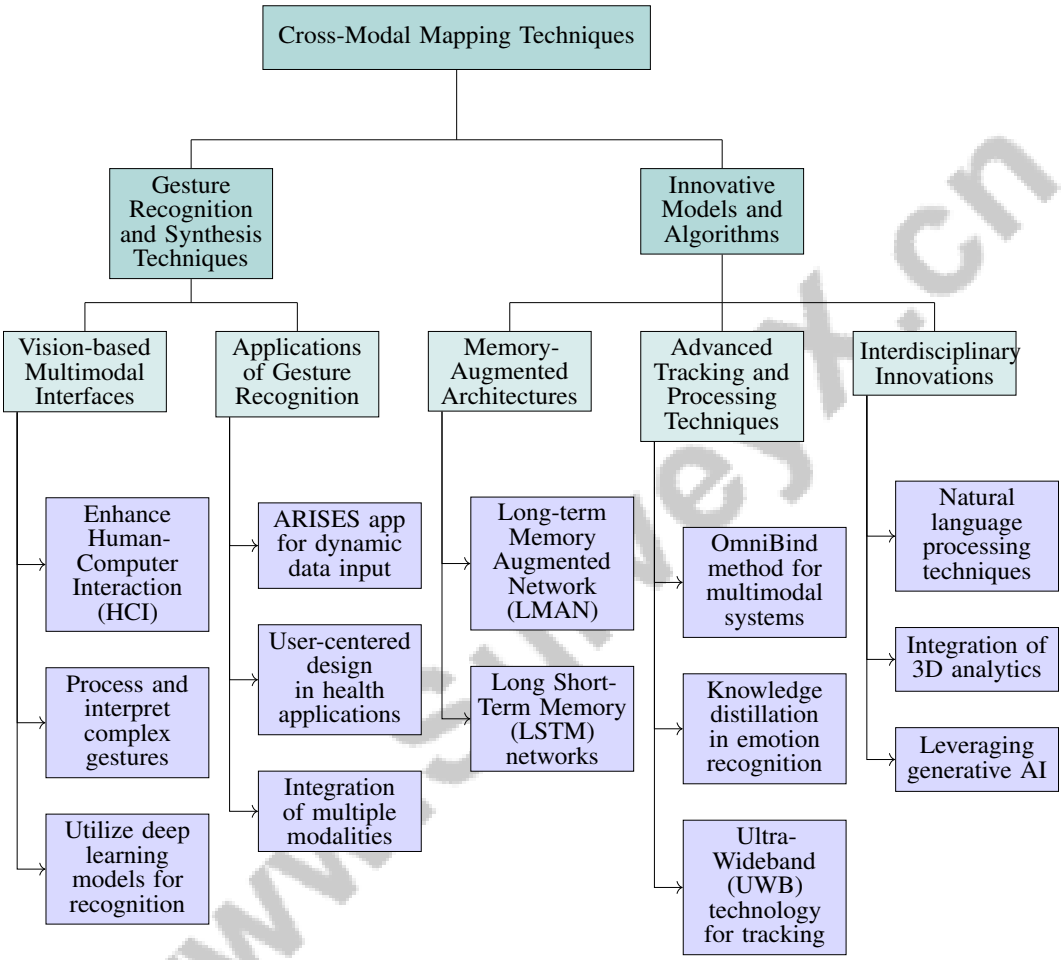


Figure 2: This figure illustrates the hierarchical structure of cross-modal mapping techniques, categorizing key advancements in gesture recognition and synthesis, innovative models and algorithms, and their applications in enhancing Human-Computer Interaction (HCI). The diagram highlights the integration of vision-based interfaces, memory-augmented architectures, and interdisciplinary innovations, emphasizing their roles in creating adaptive and user-centered interaction systems.

3 Cross-Modal Mapping Techniques

3.1 Gesture Recognition and Synthesis Techniques

Gesture recognition and synthesis are pivotal in enhancing Human-Computer Interaction (HCI) by facilitating intuitive user interactions. Vision-based Multimodal Interfaces (VMIs) exemplify advancements in processing and interpreting complex gestures, essential for dynamic applications [2]. A notable development is a real-time framework that processes raw capacitive signals to detect hand gestures by utilizing deep learning models, thereby enhancing recognition accuracy [24]. Such innovations are crucial for developing responsive interfaces that cater to diverse user needs.

The ARISES app illustrates practical applications of gesture recognition, using a hand-held interface to dynamically input data and explore relationships, thereby boosting user engagement and personalization [5]. Its use in health applications highlights gesture recognition's potential for user-centered design and improved accessibility.

Advancements in gesture recognition and synthesis emphasize the integration of multiple modalities and sophisticated modeling approaches, enabling the creation of responsive, intuitive, and accessible interaction systems. Insights from interdisciplinary research, including science fiction, inspire novel HCI design approaches [25, 26, 19].

3.2 Innovative Models and Algorithms

Method Name	Model Architecture	Integration Techniques	Application Domains
LMAN[27]	Memory Augmented Network	External Memory Queue	Virtual Environments
DL-HAR[6]	Lstm Networks	Sensor Fusion	Mobile Applications
OB[28]	Mixture-of-Expert	Weight Routing	Human-computer Interaction
KDMF[29]	Masked Training	Knowledge Distillation	Emotion Recognition
CREM[9]	Encoder-decoder Architecture	Score Fusion	Emotion Recognition
CIMRS[30]	Uwb Technology	Uwb Sensors	Immersive Theater
BPE[12]	Transformer-based Classifier	Bpe Encoding	Human-computer Interaction

Table 1: Table illustrating various innovative models and algorithms, detailing their model architectures, integration techniques, and application domains. These methods highlight advancements in cross-modal mapping, showcasing their impact on enhancing interaction systems across diverse domains.

Innovative models and algorithms are essential for advancing cross-modal mapping, enhancing the integration and synchronization of diverse data types in HCI. Table 1 provides a comprehensive overview of innovative models and algorithms, emphasizing their architectural designs, integration strategies, and application areas within the context of cross-modal mapping and human-computer interaction. The Long-term Memory Augmented Network (LMAN) employs an external memory queue to store long-term gesture features, improving classification accuracy [27]. This highlights the importance of memory-augmented architectures in strengthening gesture recognition.

In human activity recognition, Long Short-Term Memory (LSTM) networks capture temporal dependencies in sensor data, exemplifying innovative model design [6]. The OmniBind method enhances multimodal systems by introducing a weight routing strategy for dynamic weight combination of modality encoders, addressing fixed weight integration limitations [28].

Knowledge distillation and masked training techniques in multimodal emotion recognition improve system capabilities by addressing unimodal shortcomings [29]. The cross-representation model, integrating high-level speech features with low-level mel-spectrograms, demonstrates the effectiveness of combining diverse feature representations [9].

Ultra-Wideband (UWB) technology introduces a novel tracking approach in interactive systems, offering larger areas and reliable tracking of concealed props at reduced costs [30]. This showcases the potential of advanced tracking technologies to enhance HCI interactivity and reliability.

The integration of natural language processing techniques, such as byte pair encoding (BPE) for action sequence vocabulary learning from mouse and keyboard interactions, exemplifies the interdisciplinary nature of innovations in cross-modal mapping [12]. This convergence of computer science, linguistics, and cognitive science improves system functionality.

These innovative models and algorithms demonstrate the transformative potential of advanced methodologies in cross-modal mapping. By synthesizing diverse data sources, recent advancements revolutionize interaction systems to be more responsive, adaptive, and user-centered, enhancing user experiences across domains such as healthcare and mobile commerce. This evolution enables personalized multimodal interactions, leveraging rich data from social networking and emerging technologies like generative AI, while the integration of 3D analytics offers opportunities for understanding user behavior and improving design in complex digital environments [25, 31, 32, 33].

AI-generated, for reference only.

4 Wearable Multimodal Data Collection

4.1 Technologies for Multimodal Data Collection

The evolution of Human-Computer Interaction (HCI) is significantly driven by the integration of multimodal data, which enhances user interaction through diverse data streams. Advanced sensors and integrated platforms, such as the ARISES app, exemplify the potential of combining user-driven and sensor-based data for comprehensive interaction analysis [5]. Wearable devices like the Oura Ring are pivotal in bio-signal acquisition, providing critical insights into user states [8]. High-resolution thermal cameras and audio devices capture nuanced emotional and physiological responses, emphasizing the importance of visual and auditory data in interpreting user inputs [34]. The integration of high-level and low-level features enhances emotion recognition from speech and text modalities [9].

Platforms such as OmniBuds expand wearable technology's scope by monitoring vital signs like heart rate and skin temperature, broadening health monitoring capabilities [10]. Smartwatches and earbuds facilitate activity data collection essential for human activity recognition (HAR) [35]. Vision-based Multimodal Interfaces (VMIs) enhance context awareness through applications like surface sensing and haptic feedback [2]. Systems integrating capacitive sensors and sophisticated middleware exemplify current technology sophistication in real-time interface control [24].

These technologies are crucial in advancing HCI by allowing computers to process information through various modalities—speech, gestures, and visual inputs—enhancing communication efficiency and reducing errors. Recent advancements in mobile devices, sensors, and machine learning techniques are pivotal in designing systems that facilitate complex multimodal interactions, underscoring their significance in HCI's evolution [36, 37, 22, 38]. By leveraging sophisticated sensing capabilities and addressing existing challenges, these technologies foster comprehensive interaction systems, enhancing user experiences across diverse contexts.

4.2 Methods of Data Acquisition

Integrating multimodal data from wearable devices is essential for advancing HCI systems, enabling precise activity recognition and user interaction modeling. This involves collecting diverse inputs, such as audiovisual data, accelerometry, and eye-tracking, providing a holistic understanding of user behavior and context. Machine learning techniques applied to these datasets enhance system responsiveness and efficiency in real-world applications [36, 22]. Various methods capture data tailored to specific application needs.

The ActionSense dataset exemplifies comprehensive data acquisition, offering 20 unique activity labels with synchronized data streams, supporting robust model development [39]. Similarly, the dataset by Zhang et al. details normal gait and freezing of gait episodes, crucial for healthcare applications [40]. Sensors such as electromyography (EMG), accelerometers, and gyroscopes are vital for accurately detecting and modeling activities, as demonstrated by the CSL dataset. The integration of multiple sensors with onboard computation, as shown by the OmniBuds platform, enhances data acquisition accuracy [10].

Advanced sensor technologies, like the SRCSM sensor, employ an independent resistance-capacitance mechanism to decouple stimuli, enabling precise joint motion recognition [41]. The sliding window approach for preprocessing activity data ensures efficient handling of large datasets, enhancing system reliability [6]. These methods contribute to developing sophisticated HCI systems capable of delivering accurate and context-aware user experiences across applications ranging from educational analytics to healthcare monitoring [42, 43, 22, 44, 45].

4.3 Solutions and Innovations

Advancements in data collection methods within wearable multimodal systems are crucial for enhancing HCI by addressing challenges like low input bandwidth and biological signal interference [14]. Innovations focus on integrating multiple sensing modalities into compact, user-friendly devices, improving functionality in real-world conditions [46].

Notable innovations include combining multimodal sensor data with machine learning to enhance fatigue assessment beyond traditional measures [47]. This approach utilizes diverse inputs for com-

prehensive user state evaluation, crucial for adaptive HCI systems. The integration of automated and manual data collection techniques exemplifies a hybrid methodology that improves data acquisition reliability [48].

Additionally, a proposed cross-modal information retrieval approach offers simplicity and efficiency, allowing effective data integration without complex training processes [49]. These innovations address challenges related to processing and retrieving information from diverse sources, enhancing overall system efficiency.

These solutions reflect ongoing efforts to overcome data collection challenges in wearable multimodal systems. By leveraging cutting-edge sensor technologies and innovative methodologies, advancements in HCI are leading to more efficient and intuitive systems, significantly enhancing user experience and interaction quality, particularly in complex digital environments where effective design is critical [37, 26].

5 Multimodal Fusion in Human-Computer Interaction

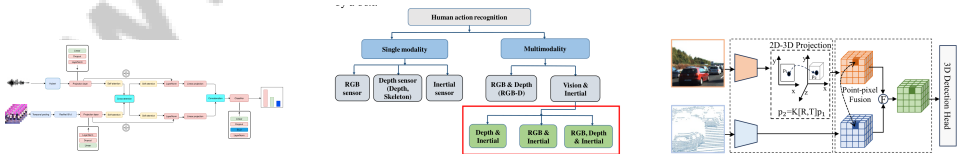
5.1 Techniques and Strategies for Multimodal Fusion

Multimodal fusion is fundamental to Human-Computer Interaction (HCI), enhancing system robustness and user experience by integrating diverse data streams. Techniques such as raw/data-level, feature-level, and score/decision-level fusion offer distinct benefits tailored to specific applications [4]. High-level feature integration, particularly in emotion recognition, combines speech and text modalities to improve model performance [9].

The ARISES app exemplifies fusion in healthcare, integrating real-time predictions with user inputs for diabetes management, underscoring the importance of HCI design in smart healthcare devices [5]. Gesture recognition combined with palmprint authentication illustrates multimodal fusion's potential to enhance security and interaction efficiency, leveraging each modality's strengths for a secure user experience [24]. High detection accuracy and low power consumption facilitate intuitive, contactless interactions [24].

Attention-based and Transformer-based approaches have surpassed traditional methods in accuracy and robustness, particularly in multimodal 3D sensing applications, enhancing data type fusion and system functionality [4]. The effectiveness of real-time hand gesture recognition systems further underscores the significance of multimodal fusion in user interaction [24].

Exploring various fusion strategies reveals the transformative potential of multimodal fusion in HCI. By integrating diverse data types, including textual data from social networks and advanced 3D representations, and employing sophisticated modeling techniques, these strategies enhance system functionality and user engagement across fields such as advertising, healthcare, and education. This comprehensive approach addresses big data and HCI complexities, fostering innovative products and services that enhance user experiences [25, 31, 26, 33].



(a) A complex neural network architecture with multiple layers and operations[50]

(b) Human Action Recognition: Single Modality vs. Multimodality[51]

(c) 3D Detection Head with 2D-3D Projection and Point-Pixel Fusion[52]

Figure 3: Examples of Techniques and Strategies for Multimodal Fusion

As shown in Figure 3, multimodal fusion is pivotal in HCI, enhancing user experiences by integrating multiple sensory modalities. The examples illustrate diverse strategies employed in multimodal fusion, each addressing unique challenges. The first example features a complex neural network architecture designed to process time series signals through a multi-layered approach, incorporating components such as linear layers, dropout, and attention mechanisms. This demonstrates how intricate network designs can effectively classify multimodal input data. The second example contrasts single modality

approaches with multimodal strategies in human action recognition, highlighting the advantages of leveraging both RGB and depth sensors for a comprehensive understanding of human actions. The third example showcases a 3D detection head utilizing 2D-3D projection and point-pixel fusion, enhancing object detection capabilities in complex environments. Collectively, these examples underscore the significance of multimodal fusion techniques in advancing HCI, offering innovative solutions for more effective and intuitive interactions [50, 51, 52].

5.2 Hybrid Interfaces and Multimodal Fusion

Hybrid interfaces play a critical role in advancing multimodal fusion in HCI, seamlessly integrating multiple sensory modalities to create natural and intuitive user experiences. These interfaces leverage various input methods—visual, auditory, and tactile—to enhance interaction fidelity and user engagement. Incorporating hybrid interfaces in multimodal systems facilitates the development of adaptive and context-aware applications, fostering a cohesive interaction environment [2].

Vision-based Multimodal Interfaces (VMIs) exemplify hybrid interfaces, combining visual and haptic feedback to enhance virtual environments without physical haptic devices [15]. This integration enables pseudo-haptic feedback, simulating touch sensations and enhancing the realism of virtual interactions, thereby improving the overall effectiveness of HCI applications.

In smart healthcare, hybrid interfaces demonstrate their value by fusing multimodal data from wearable devices, such as physiological signals and motion data, enhancing the accuracy of health monitoring and diagnosis [4]. This integration fosters the development of user-centered healthcare solutions that are both effective and intuitive, addressing the diverse needs of patients and healthcare providers.

Moreover, hybrid interfaces contribute to more secure and efficient authentication systems by combining modalities like gesture and palmprint recognition [24]. This fusion enhances security robustness and facilitates seamless user interactions, highlighting the potential of hybrid interfaces to improve system security and usability.

6 Motion Analysis and Text-Based Interaction

6.1 Role of Motion Analysis in Gesture Recognition

Motion analysis is crucial for enhancing the intuitiveness of gesture recognition in Human-Computer Interaction (HCI). By integrating motion analysis with multimodal data, systems can create adaptive interaction environments. For instance, a speech-driven selective state space model facilitates realistic gesture generation in response to spoken inputs [53]. The MAIA framework exemplifies real-time facial expression analysis, demonstrating motion analysis's role in producing non-verbal responses that heighten user engagement, especially in virtual reality and social robotics [54]. Furthermore, motion analysis enhances security and usability through mid-air gesture-based user authentication, recognizing unique gesture patterns [55]. Employing advanced motion capture techniques in gesture recognition fosters natural, secure, and effective interactions, which significantly enhance user experiences in areas like affective computing and augmented reality by improving body representation and multimodal feature integration [56, 57, 58, 59, 21].

6.2 Facial and Body Motion Analysis

Facial and body motion analysis are pivotal for interpreting user behavior and emotions, thereby advancing HCI systems. Sophisticated algorithms and sensor technologies enable precise data capture, such as facial electromyography (EMG) for detailed muscular activity assessments, crucial for indexing emotional valence [16]. Real-time emotion recognition systems utilizing high-resolution thermal cameras and audio devices extract high-level features from visual and auditory data, enhancing emotion recognition accuracy [34, 9]. Body motion analysis employs sensors like accelerometers and gyroscopes to capture detailed movement patterns, as demonstrated by the CSL dataset [60]. Advanced sensor technologies, such as the SRCSM sensor, improve joint motion recognition precision by decoupling pressure and strain stimuli [41]. These techniques enrich interaction modeling, creating more intuitive systems that significantly improve user experiences across domains like education,

healthcare, and digital interfaces, addressing specific challenges and paving the way for personalized interactions [56, 26, 61, 22, 33].

6.3 Text-Based Interaction and Multimodal Integration

Text-based interaction serves as a fundamental modality in HCI, providing a versatile communication means. Integrating text with modalities like gesture and voice enhances interaction richness and context-awareness, crucial for seamless user experiences. This integration is particularly beneficial in assistive technologies, where gesture recognition complements text input to support individuals with disabilities [62]. Natural Language Processing (NLP) techniques are essential for enabling systems to comprehend and generate human-like responses. Combining NLP with gesture recognition allows for more accurate interpretation of user intentions, resulting in context-aware responses that enhance engagement [12]. This is evident in applications where text and gesture inputs control devices or navigate virtual environments. Additionally, fusing text-based interaction with visual and auditory modalities bolsters emotion recognition systems, achieving higher accuracy by analyzing textual data alongside facial expressions and vocal cues [9]. This multimodal approach fosters a comprehensive understanding of user states, enhancing HCI system adaptability and responsiveness. By leveraging the strengths of various modalities and employing advanced processing techniques, these systems significantly enhance user experiences, offering tailored interactions across diverse applications. Addressing challenges in multimodal machine learning, strategies like novel regularization terms and pretrained deep learning models optimize decision-making processes, improving performance in multimedia analysis, human-computer interaction, and embodied AI, ultimately leading to more effective and personalized user experiences [49, 63].

7 Applications and Case Studies

This section examines the significant impacts of cross-modal mapping and wearable multimodal data across various domains, with a focus on healthcare and emotion recognition. These technologies hold transformative potential for enhancing health monitoring and emotional understanding.

7.1 Applications in Healthcare and Emotion Recognition

The integration of cross-modal mapping and wearable multimodal data has revolutionized healthcare, particularly in enhancing the accuracy and effectiveness of health monitoring systems. Wearable devices with multimodal sensors enable continuous tracking of physiological signals, which is crucial for managing chronic conditions and promoting preventive healthcare [64]. The ARISES app exemplifies this by using real-time data from wearable sensors to aid Type 1 diabetes patients in self-management, providing timely feedback and fostering engagement [5].

In emotion recognition, combining multimodal data—visual, auditory, and textual—has led to more robust systems. The Speaking Faces dataset illustrates how high-resolution thermal cameras and audio devices enhance emotion recognition by capturing nuanced expressions [34]. Incorporating high-level features further improves the interpretation of complex emotional states from speech and text [9].

Natural Language Processing (NLP) techniques underscore the interdisciplinary nature of these systems, with methods like byte pair encoding (BPE) capturing complex interaction goals for more accurate real-world interaction representations [12]. This approach increases emotion recognition accuracy and aids in developing context-aware, adaptive HCI systems.

The integration of cross-modal mapping and wearable multimodal data in healthcare and emotion recognition showcases significant transformative potential. Unsupervised multimodal representation learning advancements enhance affective computing through biosignal analysis like ECG and EDA, reducing reliance on manual feature extraction and improving emotion detection systems' generalizability across diverse datasets. Deep learning models in multimodal emotion detection outperform traditional unimodal methods, enriching the understanding of human emotional states and enhancing user experiences across various applications [65, 22, 66].

7.2 Healthcare and Public Health Monitoring

Wearable technology is invaluable in healthcare and public health monitoring, offering capabilities for continuous data collection and real-time physiological signal analysis. Multimodal data integration from wearable devices enables comprehensive health metric monitoring, crucial for early detection and intervention in public health scenarios. Wearable sensors tracking vital signs like heart rate, temperature, and movement patterns provide insights into health trends, facilitating timely public health responses [10].

Wearable technology also aids in detecting and managing infectious diseases. Advanced sensors can track viral infection symptoms, such as heart rate or temperature changes, enabling early identification and outbreak containment [8]. This capability is crucial during pandemics for informing public health strategies and mitigating disease spread.

Integrating machine learning algorithms with wearable data enhances health monitoring systems' predictive capabilities. By analyzing physiological data patterns, these systems forecast health outcomes and provide personalized disease prevention and management recommendations [6]. This approach improves individual health outcomes and contributes to broader public health objectives by identifying at-risk populations and informing targeted interventions.

7.3 Augmented and Virtual Reality

Augmented and Virtual Reality (AR and VR) technologies have transformed Human-Computer Interaction by creating immersive environments that enhance user engagement and experience. These technologies leverage cross-modal mapping and wearable multimodal data to provide realistic and interactive experiences. In AR and VR applications, integrating multimodal data streams—visual, auditory, and haptic—enhances virtual environments' realism and interactivity [2].

Vision-based Multimodal Interfaces (VMIs) in AR and VR systems highlight the potential of integrating sensory modalities to enhance context awareness and user interaction, blending virtual and real-world elements for a more intuitive experience [2]. Pseudo-haptics in VR systems further illustrate multimodal integration's potential to create realistic tactile sensations without physical haptic devices [15].

Advanced motion analysis techniques in VR environments enhance gesture recognition systems' accuracy and responsiveness. Detailed motion data capture facilitates natural interactions, essential for immersive experiences [53]. Integrating facial and body motion analysis enriches user experiences by providing real-time feedback and enhancing virtual avatars' realism [54].

7.4 Assistive Technologies and Accessibility

Assistive technologies leveraging cross-modal mapping and wearable multimodal data have significantly improved accessibility for individuals with disabilities by offering tailored solutions that enhance interaction and engagement. Integrating diverse sensory modalities—gesture and voice recognition—facilitates developing intuitive interfaces addressing users' unique needs [62].

Gesture-based interaction systems enable users to communicate and control devices through natural movements, utilizing sophisticated motion analysis techniques to accurately interpret gestures, providing alternatives to traditional input methods, and enhancing accessibility for individuals with physical impairments [55]. Kinect sensor-based gesture recognition exemplifies these technologies' potential in creating inclusive interaction environments through hands-free interfaces for users with limited mobility [62].

Integrating NLP techniques in assistive technologies enhances systems' ability to understand and respond to user commands, making communication more seamless and effective. Combining text-based interaction with modalities like voice and gesture delivers context-aware, adaptive responses, enhancing user experiences for individuals with communication challenges [12].

Developing multimodal interfaces incorporating tactile feedback further enhances accessibility by providing sensory cues aiding navigation and interaction. These interfaces leverage multiple modalities' strengths to create immersive experiences, allowing users to engage with digital environments in ways that best suit their abilities [15].

The integration of cross-modal mapping and wearable multimodal data in assistive technologies markedly improves accessibility for individuals with disabilities through innovative interaction methods, such as gaze-assisted communication, enabling users with motor or speech impairments to engage with technology effectively. Machine learning techniques applied to wearable sensor data facilitate developing systems addressing common user interaction challenges, enhancing user experience and promoting inclusivity [67, 22]. By integrating diverse sensory modalities and employing advanced processing techniques, these technologies offer personalized and effective interaction solutions, enhancing users' quality of life and independence.

8 Challenges and Future Directions

8.1 Technical and Implementation Challenges

Implementing cross-modal mapping and wearable multimodal data systems in Human-Computer Interaction (HCI) presents significant technical challenges that impact their scalability and effectiveness. Accurate segmentation and classification are critical, particularly in complex scenarios like counting repetitions in exercises with alternating movements [35]. Usability issues, such as discomfort from prolonged use of devices like the HoloLens, further hinder adoption [59]. Environmental factors, including limited detection ranges and sensitivity to variations, affect signal stability in real-time interface control [24]. Systems requiring precise anchor placement may face difficulties in high-interference environments [30].

The quality of training data is crucial for model generalization across diverse contexts, influencing the robustness of multimodal emotion recognition systems [9]. The integration of Vision-based Multimodal Interfaces (VMIs) with emerging AI technologies is underexplored, necessitating comprehensive strategies for enhanced responsiveness [2]. Additionally, integrating user needs into device design is challenging, particularly concerning safety and usability, with low adoption rates often due to usability issues [4]. The reliance on technology requires user adaptation to new interfaces, impacting the effectiveness of applications like the ARISES app [5].

Addressing these challenges is essential for advancing HCI. Improvements in data integration techniques, analytical model adaptability, and diverse dataset creation can unlock the potential of cross-modal mapping and wearable multimodal data systems. This progress will support the development of user-centered interaction technologies, particularly in educational analytics, where extracting meaningful patterns from diverse sensor data can enhance teaching practices and learning outcomes. Advanced analytics in three-dimensional representations and social data can further enrich user experiences in complex digital environments [31, 33, 22, 56].

8.2 Security, Privacy, and Ethical Concerns

The use of cross-modal mapping and wearable multimodal data in HCI raises significant security, privacy, and ethical concerns, particularly regarding sensitive data handling. The reliance on wearable devices demands robust encryption and access control to safeguard data privacy, with ongoing challenges in securing device communications and data storage [42]. Unauthorized access and data breaches threaten user privacy and trust in wearable technologies [68].

Ethical considerations include transparency and accountability in AI systems, crucial for human-centered design principles. The complexity of AI behavior and the risk of biased responses necessitate standardized evaluation methods to ensure ethical AI development and deployment. Current research often overlooks these ethical implications, highlighting the need for comprehensive frameworks addressing the moral and societal impacts of automation and AI technologies [69].

Ethical data collection and usage are paramount, particularly regarding informed consent and participant privacy. The ActionSense dataset, for instance, emphasizes transparency and respect for autonomy in data collection practices [39]. However, challenges in data interpretation consistency and reliability persist, especially with subjective modalities like voice features [70].

In virtual reality, handling sensitive data such as eye-tracking information raises additional privacy concerns, necessitating stringent ethical guidelines and security protocols [68]. These issues underscore the need for interdisciplinary collaboration to develop solutions addressing the multifaceted ethical, security, and privacy challenges in HCI [3].

Addressing these concerns is crucial for developing HCI technologies that respect user rights and foster trust. Researchers and developers should adopt comprehensive security protocols, promote transparency in AI decision-making, and emphasize ethical data handling practices. This approach ensures user needs are prioritized while mitigating AI deployment risks, leading to more intuitive and accessible systems across industries [25, 69, 71, 72].

8.3 Future Research Directions

Future research in cross-modal mapping and wearable multimodal data within HCI should explore avenues to enhance system adaptability and robustness. Developing intuitive and user-friendly smart healthcare technologies, integrating AI and machine learning for personalized healthcare, and refining user-centered design principles are key areas [4]. Expanding gesture recognition datasets through unsupervised learning could improve system adaptability and personalization [24]. Research should also enhance gesture recognition frameworks' robustness to manage complex interactions and incorporate additional communication modalities [35].

Integrating Electromyography (EMG) with other physiological measures presents promising research opportunities. Future studies should refine EMG interpretation methods and explore applications in complex social interactions to enhance user state assessments [3]. Enhancing systems like SimplyMime with additional sensors and exploring applications in virtual and augmented reality could improve interaction systems' versatility [5]. Future research should also focus on expanding methods to other cross-modal tasks, particularly in challenging environments, to enhance robustness and generalization [2]. Developing tools to measure unconscious interactions and refining interface designs could lead to more user-friendly software interfaces [12].

Expanding studies to larger populations and incorporating additional sensing modalities is essential [30]. Improving data acquisition methods, enhancing model robustness, exploring hybrid models, and leveraging transfer learning and generative models are crucial for addressing current limitations [9]. Further research should focus on enhancing robustness against varying light conditions and investigating 3D gesture recognition capabilities [59].

Expanding user research methodologies and emphasizing diverse participant recruitment are critical for addressing gaps in current studies [73]. Optimizing models like COGMEN for real-time emotion recognition and examining the impact of different hyper-parameter settings on performance could significantly advance the field [9].

Refining software development platforms, enhancing human-computer interaction methods, and exploring big data implications in wearable technology are pivotal for future advancements [24]. Refining algorithms for diverse populations and exploring real-time deployment of detection systems can improve applicability [30].

Expanding datasets with in-the-wild data to address current limitations and exploring additional multimodal tasks are essential for advancing the field [9]. The integration of higher-quality and larger-scale multimodal representations can enhance capabilities and open new applications [2]. Further enhancements in sensor accuracy, battery efficiency, and the development of applications leveraging platforms like OmniBuds are promising research directions [3].

Engaging with emerging trends and research opportunities will allow HCI to progress toward creating more effective, user-centered systems. This advancement is crucial for enhancing interaction capabilities and improving user experiences across diverse application domains, as AI and user experience design integration become increasingly vital. Addressing digital device complexities and the influence of science fiction narratives on HCI research can lead to innovative design principles and methodologies that foster greater usability, safety, and efficiency in technology [37, 71, 26, 19].

9 Conclusion

The exploration of cross-modal mapping and wearable multimodal data underscores their transformative impact on Human-Computer Interaction (HCI). By integrating diverse sensory inputs such as gestures, voice, and physiological signals, these technologies enhance interaction systems' intuitiveness and adaptability, fostering more natural user experiences. Notably, the ability of simple touch gestures to maintain low authentication error rates exemplifies the potential of straightforward

interaction techniques in improving system effectiveness. The identification of core HCI principles further emphasizes the importance of a unified design approach, which is crucial for developing cohesive and user-centered systems. Resources like the HoMG database offer significant advancements in micro-gesture recognition, particularly benefiting applications in augmented and virtual reality. Wearable devices, pivotal in health monitoring, demonstrate their utility in early disease detection, exemplified by their role in identifying COVID-19 symptoms. Additionally, advancements in eye movement recording systems contribute to seamless interactions, enhancing user experience. However, challenges remain, particularly in continuous gesture recognition and the robustness of multimodal systems. Addressing these issues requires focused research on enhancing system capabilities, especially with small datasets and real-time processing, to bolster the reliability and effectiveness of HCI technologies.

References

[1] Sayem Mohammad Siam, Jahidul Adnan Sakel, and Md. Hasanul Kabir. Human computer interaction using marker based hand gesture recognition, 2016.

[2] Yongquan Hu, Wen Hu, and Aaron Quigley. Towards enhanced context awareness with vision-based multimodal interfaces, 2024.

[3] Wei Xu. User centered design (vi): Human factors approaches for intelligent human-computer interaction, 2022.

[4] Pu Liu, Sidney Fels, Nicholas West, and Matthias Görges. Human computer interaction design for mobile devices based on a smart healthcare architecture, 2019.

[5] Robert Spence, Chukwuma Uduku, Kezhi Li, Nick Oliver, and Pantelis Georgiou. A novel hand-held interface supporting the self-management of type 1 diabetes, 2020.

[6] Seungeun Chung, Jiyoun Lim, Kyoung Ju Noh, Gague Kim, and Hyuntae Jeong. Sensor data acquisition and multimodal sensor fusion for human activity recognition using deep learning. *Sensors*, 19(7):1716, 2019.

[7] Shibo Zhang, Yaxuan Li, Shen Zhang, Farzad Shahabi, Stephen Xia, Yu Deng, and Nabil Alshurafa. Deep learning in human activity recognition with wearable sensors: A review on advances, 2022.

[8] Ashley E Mason, Frederick M Hecht, Shakti K Davis, Joseph L Natale, Wendy Hartogensis, Natalie Damaso, Kajal T Claypool, Stephan Dilchert, Subhasis Dasgupta, Shweta Purawat, et al. Detection of covid-19 using multimodal data from a wearable device: results from the first tempredict study. *Scientific reports*, 12(1):3463, 2022.

[9] Mariana Rodrigues Makiuchi, Kuniaki Uto, and Koichi Shinoda. Multimodal emotion recognition with high-level speech and text features, 2021.

[10] Alessandro Montanari, Ashok Thangarajan, Khaldoon Al-Naimi, Andrea Ferlini, Yang Liu, Ananta Narayanan Balaji, and Fahim Kawsar. Omnibuds: A sensory earable platform for advanced bio-sensing and on-device machine learning, 2024.

[11] Mihai Bâce, Sander Staal, and Andreas Bulling. How far are we from quantifying visual attention in mobile hci?, 2019.

[12] Guanhua Zhang, Matteo Bortoletto, Zhiming Hu, Lei Shi, Mihai Bâce, and Andreas Bulling. Exploring natural language processing methods for interactive behaviour modelling, 2023.

[13] Min Chen. The value of interaction in data intelligence, 2023.

[14] Kirill A. Shatilov, Dimitris Chatzopoulos, Lik-Hang Lee, and Pan Hui. Emerging natural user interfaces in mobile computing: A bottoms-up survey, 2019.

[15] Rui Xavier, José Luís Silva, Rodrigo Ventura, and Joaquim Jorge. Pseudo-haptics survey: Human-computer interaction in extended reality teleoperation, 2024.

[16] Niklas Ravaja, Benjamin Cowley, and Jari Torniainen. A short review and primer on electromyography in human computer interaction applications, 2016.

[17] Frédéric Vella, Flavien Clastres-Babou, Nadine Vigouroux, Philippe Truillet, Charline Calmels, Caroline Mercadier, Karine Gigaud, Margot Issanchou, Kristina Gourinovitch, and Anne Garaix. User centred method to design a platform to design augmentative and alternative communication assistive technologies, 2022.

[18] Jiehui Jia, Huan Zhang, and Jinhua Liang. Bridging discrete and continuous: A multimodal strategy for complex emotion detection, 2024.

[19] Philipp Jordan, Omar Mubin, Mohammad Obaid, and Paula Alexandra Silva. Exploring the referral and usage of science fiction in hci literature, 2018.

[20] Sibi Chakkaravarthy Sethuraman, Gaurav Reddy Tadkapally, Athresh Kiran, Saraju P. Mohanty, and Anitha Subramanian. Simplymime: A control at our fingertips, 2023.

[21] Junxiao Xue, Jie Wang, Xuecheng Wu, and Qian Zhang. Affective video content analysis: Decade review and new perspectives, 2024.

[22] Luis Pablo Prieto, Kshitij Sharma, Łukasz Kidzinski, María Jesús Rodríguez-Triana, and Pierre Dillenbourg. Multimodal teaching analytics: Automated extraction of orchestration graphs from wearable sensor data. *Journal of computer assisted learning*, 34(2):193–203, 2018.

[23] Jungpil Shin, Abu Saleh Musa Miah, Md. Humaun Kabir, Md. Abdur Rahim, and Abdullah Al Shiam. A methodological and structural review of hand gesture recognition across diverse data modalities, 2024.

[24] Hunmin Lee, Jaya Krishna Mandivarapu, Nahom Ogbazghi, and Yingshu Li. Real-time interface control with motion gesture recognition based on non-contact capacitive sensing, 2022.

[25] Philipp Jordan. Investigating the intersection of science fiction, human-computer interaction and computer science research, 2018.

[26] S. Thuseethan and S. Kuhanesan. Effective use of human computer interaction in digital academic supportive devices, 2015.

[27] Lizhi Zhao, Xuequan Lu, Qianyue Bao, and Meili Wang. In-place gestures classification via long-term memory augmented network, 2022.

[28] Zehan Wang, Ziang Zhang, Hang Zhang, Luping Liu, Rongjie Huang, Xize Cheng, Hengshuang Zhao, and Zhou Zhao. Omnibind: Large-scale omni multimodal representation via binding spaces, 2024.

[29] Muhammad Muaz, Nathan Paull, and Jahnavi Malagavalli. Bridging modalities: Knowledge distillation and masked training for translating multi-modal emotion recognition to uni-modal, speech-only emotion recognition, 2024.

[30] Chanwoo Lee, Kyubeom Shim, Sanggyo Seo, Gwonu Ryu, and Yongsoon Choi. Never tell the trick: Covert interactive mixed reality system for immersive storytelling, 2024.

[31] Mohamed Mostafa. Modelling and analysing behaviours and emotions via complex user interactions, 2019.

[32] J. Bieniek, M. Rahouti, and D. C. Verma. Generative ai in multimodal user interfaces: Trends, challenges, and cross-platform adaptability, 2024.

[33] Gunther Gust, Tobias Brandt, Otto Koppius, Markus Rosenfelder, and Dirk Neumann. 3d analytics: Opportunities and guidelines for information systems research, 2023.

[34] Madina Abdrakhmanova, Askat Kuzdeuov, Sheikh Jarju, Yerbolat Khassanov, Michael Lewis, and Huseyin Atakan Varol. Speakingfaces: A large-scale multimodal dataset of voice commands with visual and thermal video streams, 2021.

[35] David Strömbäck, Sangxia Huang, and Valentin Radu. Mm-fit: Multimodal deep learning for automatic exercise logging across sensing devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(4):1–22, 2020.

[36] Muhammad Zeeshan Baig and Manolya Kavakli. Multimodal systems: Taxonomy, methods, and challenges, 2020.

[37] V. Hinze-Hoare. The review and analysis of human computer interaction (hci) principles, 2007.

[38] Paul Pu Liang, Akshay Goindani, Talha Chafekar, Leena Mathur, Haofei Yu, Ruslan Salakhutdinov, and Louis-Philippe Morency. Hemm: Holistic evaluation of multimodal foundation models, 2024.

[39] Joseph DelPreto, Chao Liu, Yiyue Luo, Michael Foshey, Yunzhu Li, Antonio Torralba, Wojciech Matusik, and Daniela Rus. Actionsense: A multimodal dataset and recording framework for human activities using wearable sensors in a kitchen environment. *Advances in Neural Information Processing Systems*, 35:13800–13813, 2022.

[40] Wei Zhang, Zhuokun Yang, Hantao Li, Debin Huang, Lipeng Wang, Yanzhao Wei, Lei Zhang, Lin Ma, Huanhuan Feng, Jing Pan, et al. Multimodal data for the detection of freezing of gait in parkinson's disease. *Scientific data*, 9(1):606, 2022.

[41] Lei Wen, Meng Nie, Pengfan Chen, Yu-na Zhao, Jingcheng Shen, Chongqing Wang, Yuwei Xiong, Kuibo Yin, and Litao Sun. Wearable multimode sensor with a seamless integrated structure for recognition of different joint motion states with the assistance of a deep learning algorithm. *Microsystems & nanoengineering*, 8(1):24, 2022.

[42] Shivram Tabibu. Communications for wearable devices, 2017.

[43] Sebastian Böttcher, Solveig Vieluf, Elisa Bruno, Boney Joseph, Nino Epitashvili, Andrea Biondi, Nicolas Zabler, Martin Glasstetter, Matthias Dümpelmann, Kristof Van Laerhoven, et al. Data quality evaluation in wearable monitoring. *Scientific reports*, 12(1):21412, 2022.

[44] Po Yang, Gaoshan Bi, Jun Qi, Xulong Wang, Yun Yang, and Lida Xu. Multimodal wearable intelligence for dementia care in healthcare 4.0: A survey. *Information Systems Frontiers*, pages 1–18, 2021.

[45] He Jiang, Xin Chen, Shuwei Zhang, Xin Zhang, Weiqiang Kong, and Tao Zhang. Software for wearable devices: Challenges and opportunities, 2015.

[46] Aashish N Patel, Tzyy-Ping Jung, Terrence J Sejnowski, et al. A wearable multi-modal bio-sensing system towards real-world applications. *IEEE Transactions on Biomedical Engineering*, 66(4):1137–1147, 2018.

[47] Hongyu Luo, Pierre-Alexandre Lee, Ieuan Clay, Martin Jaggi, and Valeria De Luca. Assessment of fatigue using wearable sensors: a pilot study. *Digital biomarkers*, 4(Suppl 1):59, 2020.

[48] Ranjay Krishna, Mitchell Gordon, Li Fei-Fei, and Michael Bernstein. Visual intelligence through human interaction, 2021.

[49] Hyunjin Choi, Hyunjae Lee, Seongho Joe, and Youngjune L. Gwon. Is cross-modal information retrieval possible without training?, 2023.

[50] Lev Evtodienko. Multimodal end-to-end group emotion recognition using cross-modal attention, 2021.

[51] Sharmin Majumder and Nasser Kehtarnavaz. Vision and inertial sensing fusion for human action recognition : A review, 2020.

[52] Yinjie Lei, Zixuan Wang, Feng Chen, Guoqing Wang, Peng Wang, and Yang Yang. Recent advances in multi-modal 3d scene understanding: A comprehensive survey and evaluation, 2023.

[53] Zunnan Xu, Yukang Lin, Haonan Han, Sicheng Yang, Ronghui Li, Yachao Zhang, and Xiu Li. Mambatalk: Efficient holistic gesture synthesis with selective state space models, 2025.

[54] Dragos Costea, Alina Marcu, Cristina Lazar, and Marius Leordeanu. Maia: A real-time non-verbal chat for human-ai interaction, 2024.

[55] Wenyuan Xu, Xiaopeng Li, Jing Tian, Yujun Xiao, Xianshan Qu, Song Wang, and Xiaoyu Ji. Which one to go: Security and usability evaluation of mid-air gestures, 2018.

[56] Jack Ratcliffe, Francesco Soave, Nick Bryan-Kinns, Laurissa Tokarchuk, and Ildar Farkhatdinov. Extended reality (xr) remote research: a survey of drawbacks and opportunities, 2021.

[57] Emma Harvey, Hauke Sandhaus, Abigail Z. Jacobs, Emanuel Moss, and Mona Sloane. The cadaver in the machine: The social practices of measurement and validation in motion capture technology, 2024.

[58] Zoya Bylinskii, Laura Herman, Aaron Hertzmann, Stefanie Hutka, and Yile Zhang. Towards better user studies in computer graphics and vision, 2023.

[59] Yunlong Wang and Harald Reiterer. The impact of augmented-reality head-mounted displays on users' movement behavior: An exploratory study, 2019.

[60] Hui Liu, Yale Hartmann, and Tanja Schultz. Csl-share: A multimodal wearable sensor-based human activity dataset, 2021.

[61] Vera Prohaska and Eduardo Castelló Ferrer. Spice: Smart projection interface for cooking enhancement, 2024.

[62] Biswarup Ganguly and Amit Konar. Kinect sensor based gesture recognition for surveillance application, 2018.

[63] Sahiti Yerramilli, Jayant Sravan Tamarapalli, Jonathan Francis, and Eric Nyberg. Attribution regularization for multimodal paradigms, 2024.

[64] Kyle T. Yoshida, Joel X. Kiernan, Allison M. Okamura, and Cara M. Nunez. Exploring human response times to combinations of audio, haptic, and visual stimuli from a mobile device, 2023.

[65] Kyle Ross, Paul Hungler, and Ali Etemad. Unsupervised multi-modal representation learning for affective computing with multi-corpus wearable data. *Journal of Ambient Intelligence and Humanized Computing*, 14(4):3199–3224, 2023.

[66] Dayo Samuel Banjo, Connice Trimmingham, Niloofar Yousefi, and Nitin Agarwal. Multimodal characterization of emotion within multimedia space, 2023.

[67] Vijay Rajanna and Tracy Hammond. A gaze-assisted multimodal approach to rich and accessible human-computer interaction, 2018.

[68] Efe Bozkir. Towards everyday virtual reality through eye tracking, 2022.

[69] Carlos Toxtli. Human-centered automation, 2024.

[70] Kristian Lukander. A short review and primer on the use of human voice in human computer interaction applications, 2016.

[71] Wei Xu and Marvin Dainoff. Enabling human-centered ai: A new junction and shared journey between ai and hci communities, 2023.

[72] Mark Adkins. What should i say? – interacting with ai and natural language interfaces, 2024.

[73] Kyoko Sugisaki. Chat-bot-kit: A web-based tool to simulate text-based interactions between humans and with computers, 2019.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.