USAARL-TECH-TR--2026-07

UNITED STATES ARMY AEROMEDICAL RESEARCH LABORATORY

Real-Time Multi-Sensor Data Collection and Processing: Challenges, Opportunities, and Insights From an Expert Panel Survey

Xiaomin Yue, Christopher J. Aura, & J. Andrew Atchley

Notice

Qualified Requesters

Qualified requesters may obtain copies from the Defense Technical Information Center (DTIC), Fort Belvoir, Virginia 22060. Orders will be expedited if placed through the librarian or other person designated to request documents from DTIC.

Change of Address

Organizations receiving reports from the U.S. Army Aeromedical Research Laboratory on automatic mailing lists should confirm correct address when corresponding about laboratory reports.

Disposition

Destroy this document when it is no longer needed. Do not return it to the originator.

Disclaimer

The views, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other official documentation. Citation of trade names in this report does not constitute an official Department of the Army endorsement or approval of the use of such commercial items.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
15-01-2026

2. REPORT TYPE
Technical Report

3. DATES COVERED *(From - To)*
October 2024 - September 2025

4. TITLE AND SUBTITLE
Real-Time Multi-Sensor Data Collection and Processing: Challenges, Opportunities, and Insights From an Expert Panel Survey

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER
6.2

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHOR(S)
Yue, X.[1,2], Aura, C. J.[1], Atchley, J. A.[1,2]

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Aeromedical Research Laboratory
P.O. Box 620577
Fort Rucker, AL 36362

8. PERFORMING ORGANIZATION REPORT NUMBER
USAARL-TECH-TR--2026-07

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Development Command
Military Operational Medicine Research Program
504 Scott Street
Fort Detrick, MD 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)
USAMRDC MOMRP

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. Approved for public release: distribution is unlimited.

13. SUPPLEMENTARY NOTES
[1]U.S. Army Aeromedical Research Laboratory, [2]Chickasaw Nation Industries

14. ABSTRACT
Monitoring and evaluating aviator's cognitive states in real-time using neurophysiological and physiological signals recorded from wearable multi-sensor systems holds promise for enhancing flight safety and improving mission success. However, several key challenges must be addressed to realize this vision. These includes synchronizing signals across different modalities, implementing robust real-time data cleaning pipelines, developing effective methods for multi-sensor data fusion, and overcoming computational constraints associated with real-time processing and model inference. This report synthesizes expert insights gathered through the targeted survey, highlighting potential solutions to these challenges and outlining strategies for enhancing real-time cognitive state monitoring and performance prediction in operational cockpit settings.

15. SUBJECT TERMS
multi-sensor recording, EEG, ECG, fNIRS, synchronization

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
UNCLAS	UNCLAS	UNCLAS	SAR	36	Loraine St. Onge, PhD
					19b. TELEPHONE NUMBER *(Include area code)* 334-255-6906

This page is intentionally blank.

Summary

Monitoring and evaluating operators' cognitive states in real-time using neurophysiological and physiological signals recorded from wearable multi-sensor systems holds promise for enhancing flight safety and promoting mission success. However, several key challenges must be addressed to realize this vision. These include synchronizing signals across different modalities, implementing robust real-time data cleaning pipelines, developing effective methods for multi-sensor data fusion, and overcoming computational constraints associated with real-time processing and model inference. This report synthesizes expert insights gathered through targeted questions, highlighting potential solutions to these challenges and outlining strategies for enhancing real-time cognitive state monitoring and performance prediction in operational cockpit settings.

This page is intentionally blank.

Table of Contents

List of Figures and Tables

This page is intentionally blank.

Introduction

The U.S. Army's Future Vertical Lift (FVL) program aims to design airframes with unparalleled performance capabilities. The FVL program's innovations, from faster speeds at lower altitudes to prolonged flights and cutting-edge reconnaissance equipment, promise enhanced coordination with group forces. However, those advancements come with their own set of challenges, such as increasing operators' workload. The increased workload on operators could lead to potential mishaps, with costs that could soar upwards of $100 million annually, as past aircraft data suggests (NATO RTO-TR-HFM-162, 2012).

Monitoring in real-time an operator's cognitive state and performance allows for the timely detection of conditions in which aviators may be impaired, distracted, or experiencing elevated stress, thereby enabling intervention by mission planners, supervisors, or automated systems to prevent accidents and operational failures. Traditional methods, such as self-assessment of cognitive workload, are often unreliable due to biases and are ineffective during periods of high cognitive workload when available cognitive resources are low. These challenges underscore the need for more objective assessment tools. Physiological and neurophysiological measures, including electrocardiogram (ECG), electroencephalogram (EEG), and functional near-infrared spectroscopy (fNIRS), provide a promising avenue, as highlighted by various studies (Caldwell, 2005; Matthews & Desmond, 2002; Saxby et al., 2013; Gao et al., 2025; Li et al., 2022; Niu et al., 2025). By leveraging these measures, we can objectively assess an operator's cognitive state, a process broadly known as operator state monitoring (OSM) (which may also encompass physiological health, hypoxia, or fatigue monitoring). The long-term goal is to deploy wearable sensor systems that support real-time cognitive monitoring, inform adaptive automation, and predict operator performance to help avert potential mishaps in complex flight environments.

Monitoring human cognitive processes with multiple neurophysiological and physiological sensors offers a more comprehensive view of brain and body dynamics than any single modality alone (Li et al., 2022). Modalities such as EEG, fNIRS, ECG, eye tracking, and other measures each capture distinct but complementary aspects of cognitive states. For example, heart rate variability (HRV), which measures the time variation between consecutive heartbeats, is a widely used indicator of autonomic nervous system (ANS) function (Forte et al., 2019). The high temporal resolution of EEG provides valuable insights into dynamic changes in workload and fatigue (Diaz-Piedra et al., 2020). When combined, they can provide richer and more integrated characterization of mental states, such as attention, workload, and fatigue, than when used in isolation.

However, integrating these data streams and performing real-time OSM *in practice* presents substantial challenges. Key among them is the need to precisely *synchronize* signals operating on different time scales, the difficulty of performing *real-time data cleaning and analysis* (including quality control and adaptive machine learning [ML]) during data recordings, the development of effective *data fusion* techniques to combine features extracted from each modality, and *computational demands* of processing large volumes of high-dimensional data in real-time rather than offline. In this report, we examine each of these challenges and outline current solutions and best practices, drawing mainly on expert insights.

This report is not intended to serve as an exhaustive literature review for each of the identified challenges, as any one of them could warrant a dedicated review on its own (e.g., Duffy & Feltman, 2023; Vogl et al., 2023). Given the number of the topics addressed, a full literature review for each would exceed the scope of this report and potentially obscure its practical focus. Instead, we place greater emphasis on expert perspectives gathered in response to targeted questions designed to address the key issues, while selectively referencing relevant literature where it directly informs key issues. This approach aims to guide the development of practical tools and models suitable for real-world OSM applications.

To inform this report, specialized subject matter experts in the research community were asked to participate in a virtual expert panel. Specifically, highly respected experts were contacted through email and asked to provide information on best methodological practices. We include a list of the experts we contacted to generate this report (Table 1), along with the specific questions asked (Table 2). To protect privacy, individual names in Table 1 have been replaced with alphabetical letters. All experts consented to publication of their affiliation and expertise as written. Full, unedited responses are included in Appendix B, organized by questions.

It is worth noting that many of the strategies discussed throughout this report to address those key issues may not appear as the main focus of journal articles. Instead, they often exist in the background of successful implementation, shared informally among research teams or mentioned briefly in method sections of articles. However, these strategies are critical for ensuring that OSM applications can be safely and effectively deployed in real-world flight environments.

This space is intentionally blank.

Table 1. Experts Contacted for the Report

Name	Affliation	Expertise
Dr. A	Department of Psychological and Brain Science Drexel University USA	Multi-sensor recording, neuroimaging, EEG, fNIRS
Dr. B	Donders Instititue for Brain and Behavior, Cognition, and Behaviour Radboud University The Netherlands	EEG, skin conductance, eye movements, and multi-modal recording
Dr. C	Meta Reality Labs (Paris) France	EEG, MEG, co-creator of MNE-python, spearheads GPU-accelerated real-time source imaging and BCI pipelines
Dr. D	Berlin Mobile Brain/Body Imaging Lab, Technische University Germany	Pioneer of MoBI + virtual reality (VR) + motion-capture fusion, high-density EEG synchronized to kinematics
Dr. E	Center for Cognitive Neuroimaing, Radboud University Netherlands	Founder of FieldTrip toolbox, real-time EEG/MEG streaming and analysis
Dr. F	Department of Engineering Technology University of Houston USA	Biomedical optics, optical bioimaging, wearable health technologies, EEG, fNIRS
Dr. G	Simpson Institute for Bioelectronics Northwestern University USA	Bio-integrated electronics, soft materials, and flexible biomedical devices, microfluidics
Dr. H	U.S. Army Research Laboratory USA	Visual neuroscience and eye tracking, EEG and neuroimaging, neuroergonomics and human-artificial intelligence (AI) teaming

Note. MEG = magnetoencephalography, MNE = MEG+EEG analysis and visualization, GPU = graphic processing unit, BCI = brain-computer interface, MoBI = mobile brain/body imaging.

Table 2. Questions for the Experts

Number	Questions
1	What level of inter-stream temporal error is still acceptable for your analyses?
2	If you have unlimited resources, what would be your “gold-standard” synchronization setup look like? Why isn’t it practical today?
3	Which artifacts are still hardest to suppress in real-time without post-hoc processing?
4	Do you trust real-time quality metrics? Why?
5	What is the biggest difference between an algorithm that performs well offline and one that survives on-device deployment?
6	How do you validate a closed-loop ML model that adapts on the fly?
7	How do you handle missing packets or corrupt segments in a long recording when the study cannot stop?
8	What is the best method you’ve seen for ground-truthing multi-modal physiological streams in the wild?
9	What is the most overlooked privacy, security risk when streaming raw physiological data wirelessly?
10	Looking five years ahead, which current research question will feel “resolved” and which challenge will still be with us?

Synchronization Across Modalities

When combining signals like EEG (sampling at 200–1000 hertz [Hz]) with slower signals like fNIRS or peripheral physiology (often 10–100 Hz), precise time synchronization is essential (Sun et al., 2021; Uchitel et al., 2021). All sensors must be referenced to a common timeline so that cognitive events (e.g., a stimulus onset or a blink artifact) align correctly across data streams. Even slight misalignments can lead to spurious results. For example, an EEG peak might be paired with the wrong heartbeat or stimulus if clocks drift. As Xiao et al. (2022) emphasized, accurate timestamp matching is the foundation for multi-modal analyses and for validating new sensors against gold-standard devices. Without synchronization, accurate interpretation of relationships between modalities is compromised. Thus, robust synchronization is a prerequisite for any multi-sensor cognitive study and OSM.

Temporal Resolution and Acceptable Error

Different applications (such as neurofeedback, fatigue detection, BCI, or OSM) tolerate different amounts of timing error between streams. Many experts maintain that high temporal resolution brain signals like EEG or MEG should be synced within only a few milliseconds. For instance, Dr. E suggested that around 5 milliseconds (ms) jitter is the maximum acceptable timing error for correlating EEG/MEG with other streams. Similarly, Dr. C and others advocate keeping inter-stream differences below ~10 ms. In experiments examining fast event-related potentials or stimuli that depend on precise timing (e.g., gaze-contingent events), even 10–20 ms offsets could blur or drown out the neural effects of interest. On the other hand, slower physiological trends can forgive a bit more lag. For example, fNIRS signals change over seconds, so misalignment on the order of hundreds of milliseconds might be acceptable for some analyses, according to Dr. F. Likewise, aggregated cognitive state metrics (which often average signals over hundreds of ms) might tolerate on the order of ±20 ms timing uncertainty (Dr. A). The general consensus is to aim for sub-10 ms synchronization error for neuropsychological signals and try to keep even slower modalities within a few tens of milliseconds alignment. In specialized cases like aligning EEG with discrete event triggers, sub-millisecond precision is ideal. These tolerances inform the required rigor of synchronization solutions.

Hardware Synchronization Methods

Hardware-based synchronization remains one of the most reliable approaches for aligning multi-modal physiological recordings, though implementation varies in precision, cost, and scalability. A commonly used method involves sending timing pulses or event markers to all devices through physical trigger lines. For example, a transistor-transistor logic (TTL) pulse can be delivered simultaneously to EEG, fNIRS, and other systems to mark shared events (e.g., Shin et al., 2018). This type of synchronization aligns data streams relative to discrete events, such as stimulus onset or task transitions, but does not provide continuous alignment at the level of individual time points. In controlled laboratory environments with physically connected equipment, this approach is low-cost, relatively simple, and sufficiently accurate for many experimental designs having well-defined and temporally distinguishable events.

A more precise synchronization approach involves the use of a shared clock or centralized acquisition system to timestamp data across modalities. In such systems, each physiological signal must be amplified individually, as EEG, ECG, fNIRS, and other modalities require signal-specific amplification and filtering. The amplified signals are then digitized and timestamped with a unified acquisition platform that applies a common timing reference. For example, recent EEG and fNIRS systems have taken advantage of these configurations to achieve precisely synchronized recordings (e.g., Gao et al., 2025).

While this centralized setup enables precise synchronization at the sample level, it introduces significant practical limitations. Each sensor requires its own amplification circuit, and scaling to multiple sensors increases system complexity, cost, and power consumption. For instance, integrating ten physiological sensors would require ten separate amplification circuits, which makes the approach cumbersome and difficult to implement in wearable or mobile environments.

To overcome these limitations, modern synchronization solutions employ distributed timing protocols over network infrastructure. One such approach involves the use of the precision time protocol (PTP), where a grandmaster clock is transmitted via Ethernet or optical fiber to each recording device. This can be combined with a synchronized global positioning system (GPS) pulse-per-second signal, allowing each device to timestamp data relative to a shared global reference. This approach enables continuous and highly precise synchronization across devices, often achieving sub-millisecond or even microsecond alignment (e.g., Lee et al., 2019). According to Dr. A, such systems represent a gold standard for synchronization by effectively eliminating clock drift. In addition, network-based synchronization offers greater scalability and flexibility, making it particularly well-suited for distributed, mobile, or wearable sensor networks where central wiring is not feasible.

Although the event-based synchronization through trigger pulses offers a simple and cost-effective method for aligning discrete events, it does not ensure precise time alignment across data streams. This limitation can be particularly problematic for OSM applications. In real-world operational environments, operators often engage in complex, continuous tasks that lack clearly defined start and stop points. Moreover, they may perform multiple tasks simultaneously, making the event-based synchronization insufficient for capturing the full temporal dynamics of cognitive and physiological processes. Centralized acquisition systems provide higher temporal precision but are limited by hardware complexity and poor scalability. In contrast, a network-based synchronization approach supports precise and continuous alignment across multiple sensors while maintaining flexibility, making them ideal for real-world multi-modal data collection.

Software and Network-Based Synchronization

An increasingly popular solution is to synchronize clocks in software over a network. The Lab Streaming Layer (LSL) (Kothe et al., 2024) is one prominent open-source framework designed for this purpose. LSL establishes a common time base across devices by continuously aligning their local clocks, typically by exchanging timestamps and compensating for drift every few seconds. All data samples from each device are tagged with LSL timestamps that reflect a globally synchronized clock, with known offset and drift for each stream. This allows disparate data streams (EEG, ECG, fNIRS, motion sensors, etc.) to be recorded together in a unified timeline, often saved in a standard format like extensible data format (XDF).

Blum et al. (2021) demonstrated that LSL-based synchronization can achieve temporal alignment virtually equivalent to traditional wired methods, even in fully mobile setups. In their tests, multiple Android phones running LSL maintained synchronization within a few milliseconds while streaming data from multiple sensors. A key advantage of software-based synchronization is its flexibility. Devices do not require physical trigger lines or shared hardware. As such, any device capable of connecting to a local network, either wirelessly or through a wired connection, can join the synchronized data pool.

However, software-based synchronization is not without limitations. Network latency and clock drift still exist and must be corrected frequently to maintain timing accuracy. LSL addresses this by defaulting to clock realignment every 5 seconds, and by logging clock offset correction, which can be used during post-processing to adjust for any residual timing discrepancies. Other network-based synchronization approaches include Network Time Protocol (NTP) or PTP on supported devices, but these typically need custom implementation on each device.

Data security is a consideration when using software-based synchronization over wireless connections. Researchers must be mindful of potential vulnerabilities when streaming physiological data. Wi-Fi signals can travel long distances, and if data streams are not encrypted, sensitive information about participants’ cognitive states or biometric signature may be exposed, as several experts have pointed out. In adversarial contexts, such information could be exploited to jeopardize the success of the mission. Therefore, in real-world operational environments, Bluetooth connections, which have shorter transmission ranges and lower risk of remote interception, may offer a more secure option for physiological data streaming.

In practice, many laboratories employ a hybrid approach to synchronization. A common approach involves sending a one-time manual synchronization trigger, such as a flash of light visible to all sensors or a keypress event, to align the initial start times across devices. Following this, a software framework, such as LSL or FieldTrip developed and maintained by Dr. E, is used to continuously adjust for clock drift during recording. Drift correction is critical because even devices that begin synchronized will slowly diverge over time due to clock frequency differences.

Software agents that periodically measure clock offsets, as LSL does, or periodic re-triggering can correct for this drift. The acceptable interval for re-synchronization depends on how fast drift accumulates. Some high-end systems drift only a few microseconds per minute, whereas consumer devices might drift several milliseconds per minute. Experts recommend limiting drift to within a few milliseconds per minute for neural data streams. For example, Dr. A suggested maintaining drift at or below 5 ms per minute for EEG/fNIRS. These considerations underscore that synchronization is not a one-time operation but an ongoing process that must be maintained throughout the entire data collection period.

Emerging and Future Solutions

With unlimited resources, researchers envision synchronization setups that are both fully wireless and highly precise. Several experts foresee a future where all wearable sensors share a common wireless clock or timestamp broadcast. For instance, Dr. D suggested that an ideal, though currently impractical, solution would involve physically connecting all sensors to one central clock or computer. While such a configuration would ensure perfect synchronization, it would also severely restrict natural movement and mobility during data collection.

Other experts, such as Dr. G and Dr. A, highlight the potential of next-generation wireless technologies (such as advanced Bluetooth protocols or ultra-wideband [UWB]) to achieve sub-millisecond synchronization without the need for physical connections. Indeed, low-jitter, multi-sensor time synchronization is widely viewed as a challenge on the verge of being resolved, with prototypes already demonstrating sub-microsecond (<1 μs) accuracy for wearable devices (Dr. A, looking ahead five years). Advances in miniaturization and the integration of dedicated clock synchronization chips into wearables are expected to further enhance timing accuracy.

In the meantime, practical solutions continue to rely on frameworks such as LSL and creative hybrid approaches that combine hardware triggers, shared video or inertial measurement unit (IMU) references, and software-based clock alignment. These methods remain essential tools for ensuring all data streams are temporally aligned.

To complement expert perspectives, we also reviewed the current literature to summarize methods commonly used in multi-sensor studies, as presented in Table 3.

Table 3. Summary of Synchronization Methods in Current Multi-Sensor Studies

Authors/Year	Sensors	Synchronization Method
Al-Shargie et al. (2016)	EEG and fNIRS	Hardware trigger method
Cao et al. (2021)	EEG, EOG, IMU, and pulse oximetor	Hardware using analog-to-digital converter (ADC)
Deligani et al. (2021)	EEG and fNIRS	Software timestamp
Gao et al. (2025)	EEG and fNIRS	Hardware trigger method
Lin et al. (2020)	EEG and fNIRS	Post-hoc alignment
Shin et al. (2018)	EEG and NIRS	Hardware trigger over a parallel port
Smith et al. (2023)	EEG, ECG, EDA, and RSP, eyetracking, and fNIRS	Hardware and software
Su et al. (2023)	EEG and fNIRS	Post-hoc alignment

Online (Real-Time) Data Analysis

Collecting multi-modal data is only the first step; interpreting and acting on it in real-time presents a far greater challenge. This section examines the difficulties in cleaning and validating physiological signals as they are being acquired, as well as the deployment of machine-learning models during ongoing experiments. We also discuss strategies researchers use to ensure data quality and analytic reliability in <u>online</u> settings, meaning during live data collection, as opposed to <u>offline</u> settings, where analysis is performed after data collection is completed. Ensuring real-time signal integrity and model performance is essential for close-loop cognitive experiments and neurofeedback applications.

Artifacts and Noise in Real-Time

Neurophysiological signals are inherently noisy and prone to contamination by various artifacts (e.g., Nunez & Srinivasan, 2006). Removing artifacts becomes significantly more difficult without the benefit of offline processing. Motion artifacts are a prime example. When a participant moves suddenly, EEG electrodes may shift or an fNIRS optode may momentarily lose contact, causing large signal disturbances. According to our expert panel, such movement-related artifacts remain the hardest to suppress in real-time. Dr. E noted that if a sensor loses touch with the skin even briefly, the resulting data loss or spike is very challenging to correct in real-time. Dr. D similarly highlighted mechanical artifacts (like cable sway or electrode impedance fluctuations due to movement) as persistent issues that current real-time systems struggle to manage. Even environmental noise can pose a serious issue in mobile or field setups. As Dr. B observed, efforts to increase ecological validity by conducting experiments out of the laboratory often introduce interference in wearable EEG recordings. Such noise may raise the "floor" of the signals, making it difficult for real-time algorithms to detect subtle cognitive effects.

Online systems typically rely on basic filtering and artifact mitigation. These include high-pass filtering to remove slow drift, notch filtering to remove electrical noise, or incorporating accelerometer data to partially correct for movement-induced artifacts. While these methods are effective for handling predictable and stationary artifacts (e.g., removing a 60 Hz powerline noise, or regressing out linear motion trends), they cannot match the sophistication of offline methods. Techniques such as independent component analysis (ICA) or artifact subspace reconstruction (ASR, Mullen et al., 2013) require access to the full dataset and cannot be executed in real-time. As Dr. A pointed out, certain artifacts can only be cleanly separated with post-hoc methods (e.g., ICA or general linear modeling) that are not feasible to run in real-time. For example, when EEG signals are contaminated by muscle activity during speech or chewing, real-time systems might apply a band-pass filter to attenuate the high-frequency EEG noise, but they often cannot fully isolate the neural signal of interest until an offline ICA is performed to remove those muscular artifacts. Similarly, fNIRS signals affected by motion-coupled blood flow changes might be partly corrected for by an online motion filter, but true separation of cerebral versus scalp blood signals usually requires an elaborate offline analysis.

In summary, current real-time artifact suppression methods are limited in their ability to handle complex or transient artifacts. Sudden or episodic movement continues to present a significant obstacle. Ongoing research in this area includes developing adaptive filtering algorithms and training machine learning models to detect and remove artifacts during data acquisition. However, robust and generalizable real-time solutions remain an area of active investigation.

Real-Time Quality Control (QC) Metrics

One way to manage data quality during an experiment is to monitor signal quality metrics in real-time. Many modern acquisition systems provide continuous feedback on indicators such as electrode impedance for EEG, signal-to-noise ratio, or statistical measures like kurtosis that reflect signal integrity. These metrics can alert the researcher to potential problems, such as a detached electrode or a sensor saturation. A critical question, however, is whether these real-time quality metrics can be trusted. Expert opinions on this issue were divided.

Dr. E emphasized that real-time quality checks are only useful if they lead to an immediate corrective action. For example, if electrode impedance rises above an acceptable threshold, the experimenter can pause to reattach an electrode. In this sense, such metrics serve as feedback to enable timely intervention. He also described scenarios in which real-time monitoring may reveal problems that are not apparent in the recorded data. For instance, an eye tracker might provide a live video feed that reveals poor calibration or tracking performance before it becomes evident in the saved gaze coordinates. In such cases, real-time monitoring may detect issues that offline review of the reduced data might miss.

However, several experts cautioned against over-reliance on these metrics. Dr. D questioned the robustness of real-time metrics, particularly under variable movement conditions. Dr. B pointed out that although it is theoretically possible to develop advanced real-time metrics that differentiate among environmental noise, device-related noise, and physiological artifacts, in practice many commercial systems provide only simplistic outputs. For example, an EEG amplifier might report only impedance values, which she finds less reliable than her own visual inspection of the signals. Dr. F acknowledged that many signal quality metrics are available and generally sufficient for a quick assessment. Nevertheless, he emphasized that post-hoc analyses remain essential for interpreting data quality in context and for performing reliable artifact rejection. Similarly, Dr. A described real-time indices as early warning signals rather than definitive assessments. These metrics are helpful for detecting obvious failures, such as a disconnected lead or sensor saturation, but they are limited in their ability to detect subtle changes or non-stationarities in the data. For this reason, his laboratory always stores raw data and conducts comprehensive offline diagnostics before drawing conclusions.

In general, real-time quality control (QC) is useful for flagging major issues and, in some cases, for triggering an adaptive system response, such as adjusting a stimulus or notifying the participant to stay still, but it cannot replace the depth of offline quality analysis. For truly critical applications, some experts, including Dr. G, recommended that one should average or integrate quality measures over longer windows (tens of seconds) to get stable estimates. Attempting to update quality assessments every second or less often results in noisy and unreliable metrics. While real-time QC monitoring allows for on-the-spot adjustments and helps researchers stay informed about the ongoing quality of the data being collected, they remain cautious and rely on thorough offline review to ensure final data integrity.

Validating Online Machine Learning Performance

A growing number of cognitive experiments now employ closed-loop paradigms, in which data streams are recorded in controlled environments and analyzed in real-time to adapt the stimulus or provide feedback (e.g., Chen & Ziegler, 2025). A key challenge in these experiments is ensuring that the ML models driving these closed-loop systems maintain accurate performance throughout the session. While validating a model offline using cross-validation is well established, real-time deployment introduces new complexities. Once a model is operating online, particularly if it is adapting continuously by learning from incoming data during the experiment, questions arise regarding how to assess its ongoing performance and prevent issues such as model drift or the delivery of inaccurate feedback.

The expert discussions have proposed several solutions to this challenge. Dr. A outlined a multi-faceted approach to verifying the reliability and performance of closed-loop models:

- **Shadow mode logging** involves running the adaptive algorithm in parallel with a standard, trusted experiment control system without the model to influence the live experiment. The model generates predictions or decisions in real-time and logs them, but the participant experiences a predetermined, fixed sequence of events. After the session, the model’s logged decisions can be compared to ground truth labels or expected outcomes. This approach allows researchers to evaluate how the model would have performed in a real-time setting without risking the validity of the live experiment. For example, in a closed-loop attention task, the adaptive model might determine the optimal moment to present a target based on EEG activity. In shadow mode, the targets are instead presented at fixed intervals, while the model's predicted presentation times are recorded for post-hoc evaluation.

- **Replay benchmarking** refers to feeding previously recorded data back through the model, or through different versions of the model, to audit its behavior. After a live session, the multi-modal data streams can be replayed through the model offline, enabling retrospective evaluation of performance. This can be particularly useful for adaptive models that update themselves over time. By replaying the same data through the model at different checkpoints, researchers can examine whether the model's behavior changed and assess whether any drift or performance degradation occurred. For instance, by comparing early and late predictions on the same input data, one can determine whether the model adapted appropriately or overfit to transient features in the training data.

- **Interleaved lock-in trials** provide another strategy to monitor model accuracy during an experiment. These trials involve periodically presenting stimuli with predictable responses, such as an oddball stimulus that is expected to elicit a characteristic neural signature. By interspersing these calibration events at regular intervals within the main task flow, researchers can assess whether the model continues to respond correctly to well-understood inputs. Because the expected outcomes of these trials are known, any significant deviations from the correct detections may indicate that the model's performance is degrading. This can serve as a signal to trigger model recalibration, reinitialization, or manual review.

Using these approaches, researchers can increase their confidence in the performance and reliability of online models. Additional recommendations include the importance of thoroughly testing the entire closed-loop system in controlled settings before deploying it in more variable or natural environments. Dr. G advised beginning validation in a tightly controlled environment, and then gradually transitioning to real-world use. This stepwise approach ensures that the model's performance remains robust as experimental conditions become noisier or less predictable.

Dr. C offered a complementary perspective, emphasizing the value of behavioral validation. If the objective of the closed-loop system is to enhance user performance, such as improving reaction times or task accuracy, then one direct measure of model effectiveness is whether those improvements are observed in practice. In this context, model validation is not limited to algorithmic metrics but also include measurable improvements in participant's behavior.

Ultimately, a model that performs well offline may still fail when deployed in a live context unless it can handle unpredictable live data streams and operate stably without reliance on offline fine-tuning. The strategies discussed above provide a practical toolkit for maintaining scientific validity and experimental reliability when integrating ML into closed-loop cognitive studies or OSM.

Data Fusion and Integration

Once data from multiple sensors are synchronized and collected (ideally with minimal noise), the next challenge is how to combine these multi-modal data to produce meaningful insights. Data fusion can occur at various stages of the processing pipeline. Researchers might choose to merge raw signals, extract features independently from each modality and then combine them or conduct separate analyses and integrate the results at the decision level. Achieving effective fusion, in which the combined output provides more information than any individual modality, remains an active area of research (e.g., John et al., 2024).

The figure illustrates a hybrid EEG–fNIRS system designed for cognitive state monitoring. In this configuration, EEG and fNIRS signals are recorded concurrently to capture complementary neural information. Each data stream undergoes modality-specific preprocessing and feature extraction before the features are fused into a unified dataset for classification or inference. EEG contributes features with high temporal resolution, capturing rapid neural oscillations or event-related potentials (e.g., Cao et al., 2021), while fNIRS contributes features with high spatial resolution, reflecting localized hemodynamic changes (e.g., Li et al., 2017). The goal of this integration is to leverage the strengths of each modality to achieve better accuracy or robustness in detecting cognitive states than would not be possible using either modality alone. Multi-modal fusion has the potential to overcome the inherent limitations of single-modality systems, leading to more precise and reliable interpretation of cognitive states. Successful multi-modal fusion strategies have been shown to improve classification accuracy in mental workload and stress detection (e.g., Deligani et al., 2021). In such cases, the fused system not only captures a broader range of relevant information but also increases resilience to noise or artifacts present in any single input stream.

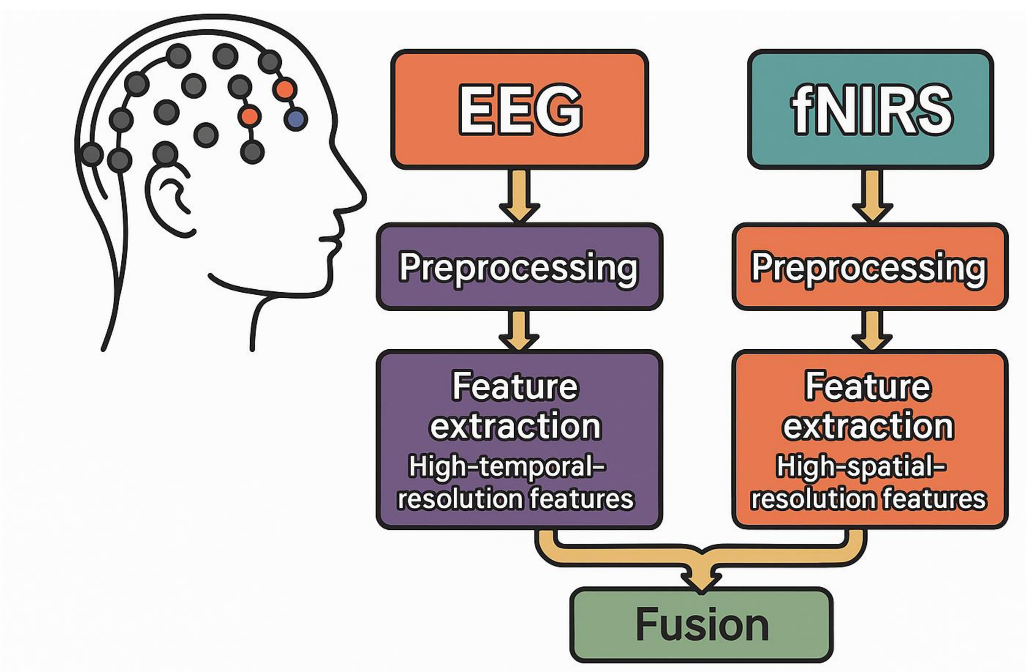

Figure 1. Example of a hybrid EEG-fNIRS setup.

Fusion Strategies (Online and Offline)

Several approaches have been developed to integrate multi-modal sensor data for cognitive state classification, each differing in where and how the information are combined (John et al., 2024). These can be broadly categorized into feature-level (early) fusion, decision-level (late) fusion, and model-level (intermediate) fusion, each with distinct assumptions, advantages, and implementation challenges.

Feature-level fusion refers to the direct concatenation or combination of extracted features from each modality into a single composite feature vector that serves as input to a downstream ML model (Niu et al., 2025; Borghini et al., 2020). For example, one might compute band-power features from EEG and HRV metrics from ECG, and then concatenate these into a unified vector to train a classifier that predicts cognitive workload or overload. This approach retains a rich representation of the multi-modal information and allows the model to learn potential cross-modal relationships. However, it requires that features from each modality be well-aligned in time, demanding precise synchronization between data streams. Moreover, the resulting high-dimensional input space may increase the risk of overfitting and impose computational burdens, particularly when large numbers of features are involved, or training data are limited (Sams et al., 2023).

In contrast, decision-level fusion involves processing each modality independently using separate models or classifiers. The outputs of these models, often probabilities or class predictions, are then integrated at a higher-level using methods such as weighted voting, rule-based logic, or ensemble learning techniques (Xefteris et al., 2023). This approach is particularly advantageous when different modalities have asynchronous sampling rates, variable reliability, or are prone to missing data. Because each modality's processing pipeline operates independently, decision-level fusion is well-suited for real-time applications and modular system design. However, this strategy may miss complex cross-modal interactions that occur at earlier processing stages (John et al., 2024).

Model-level fusion, also referred to as intermediate or hybrid fusion, occupies a conceptual space between early and late fusion. Here, each modality is processed through a dedicated subnetwork (e.g., a neural network branch) that learns modality-specific representations, which are subsequently merged within the model (typically at hidden or intermediate layers) (e.g., Chen et al., 2022). This enables the system to learn rich, joint representations that capture both intra- and inter-modal relationships. Classical statistical approaches such as canonical correlation analysis (CCA) and joint ICA exemplify this fusion strategy by identifying latent components that covary across modalities. More recently, deep learning architectures have emerged that incorporate model-level fusion via attention mechanisms, gating modules, or shared latent spaces (e.g., Zou et al., 2024). One prominent example is the Temporal Fusion Transformer (TFT), which accepts feature-level inputs but performs learned fusion internally, using a combination of attention and gating layers to selectively integrate information across modalities and time. Although inputs to the TFT may resemble those used in early fusion, its internal architecture qualifies it as a model-level fusion method, as it learns to emphasize cross-modal dependencies during training (Lim et al., 2021).

Collectively, these fusion strategies offer complementary advantages. Feature-level fusion is straightforward to implement and can exploit all available data simultaneously but requires synchronization and dimensionality control. Decision-level fusion offers robustness and modularity, especially when modalities vary in quality or temporal characteristics. Model-level fusion, while more complex to implement, offers a principled and flexible way to learn high-level multi-modal abstractions, potentially improving model generalizability and interpretability.

From a workflow perspective, offline data fusion performed after data collection enables more complex and flexible operations. Researchers can experiment with tuning parameters, retrospectively aligning data streams or addressing missing data (e.g., Gao et al., 2025; Deligani et al., 2021; Li et al., 2017) to get better results. Offline analysis also allows comparisons between fusion at the feature level and at the decision level, enabling the selection of the approach that yields better validation accuracy.

In contrast, online fusion during live data acquisition must be computationally efficient and suitable for real-time processing. This often involves predefined computations that integrate data streams as they are received. For example, a real-time cognitive load index might be calculated as a weighted combination of normalized EEG beta power and pupil diameter, using weights determined in prior offline training. Some online systems can implement heuristic rules. For instance, an attention monitoring system may require both an EEG marker and an eye gaze measure to indicate a lapse before flagging users. This approach, a form of decision level fusion, can improve specificity by reducing false positives.

A central challenge in multi-modal integration lies in the vast design space. Researchers must determine both how and when to fuse information, and these decisions often depend on domain knowledge, task requirements, and empirical testing (e.g., Li et al., 2022).

Evaluating Fusion Approaches

It is essential to evaluate whether a multi-modal combination is genuinely effective. Simply recording multiple data streams is not inherently valuable unless their integration yields new insights or leads to improved performance on a target task. Dr. D noted that many studies described as multi-modal still analyze each modality independently. Truly integrated analyses, where modalities inform and complement one another, remain relatively rare. The overarching goal is to achieve effective synergy. For example, physiological signals such as heart rate or skin conductance can be used to contextualize neural activity measured with EEG, or neural signals can be used to interpret physiological responses.

One straightforward evaluation metric is performance improvement. A critical question is whether a model trained on combined data predicts cognitive outcomes, such as error rates or mental workload, more accurately than models based on single modalities. Many studies report such gains (Lin et al., 2017; Aghajani et al., 2017; Deligani et al., 2021; Chu et al., 2022). For instance, in mental workload classification, EEG data may offer moderate predictive value, and fNIRS may also perform moderately well, but their combination can result in significantly higher accuracy (Lin et al., 2017).

Beyond accuracy, multi-modal systems offer improved robustness. These systems are more resilient to noise or data loss in any one channel. If EEG becomes excessively noisy during a task, complementary signals from ocular or cardiac sensors may still provide sufficient information to infer cognitive state, allowing the fused output to remain reliable. As Dr. G observed, redundancy across sensors can also serve as a consistency check. For example, heart rate can be measured simultaneously using ECG, photoplethysmography (PPG), and seismocardiography (SCG), such that if one modality fails, the others can provide backup measurements.

A specialized application of multi-modal integration involves using one modality to clean or validate another. Dr. F offered a forward-looking example, combining motion data from wearable IMUs with neuroimaging data to detect and correct motion related artifacts. In this scenario, the IMU can be used to identify periods of movement, enabling algorithms to remove or adjust brain signals that are likely contaminated. This form of cross modality support offers a pragmatic fusion strategy, where one sensor enhances the data quality of another. Although such corrections are currently performed manually, for instance by visually rejecting EEG segments that coincide with IMU detected movement, future systems may automate this process in real-time.

Effective multi-modal data integration presents both a valuable opportunity and a substantial challenge. It holds the promise of offering a more complete understanding of cognitive processes. However, this potential can only be realized through careful methodological design. Best practices emerging from recent research include ensuring precise temporal synchronization to avoid alignment errors during fusion, reducing each modality to its most informative features to minimize computational burden, and selecting fusion strategies that match the specific research or operational objective, whether it involves real-time monitoring or offline analysis, prediction or exploratory discovery. As Dr. D noted, even within the next five years, truly integrated multi-modal fusion will likely remain an active area of development. Continued progress is expected across both statistical and AI methods that aim to process and learn from complex multi-modal data streams in a unified framework.

Computational Demands: Online vs. Offline Processing

A critical yet often underappreciated challenge in real-time cognitive state monitoring is the computational burden associated with online processing of multi-modal data. Unlike offline analysis, which benefits from unlimited time and computational resources, real-time systems must operate under strict latency and processing constraints. This section compares these two modes and outlines the implications for algorithm design and deployment in live cognitive monitoring systems and wearable devices.

Latency and Throughput Constraints

Offline analysis allows researchers to record hours of data and process it at leisure on high performance workstations or computing clusters. For example, an algorithm that requires ten hours to analyze a one-hour dataset is acceptable in offline contexts. In contrast, online processing must keep pace with incoming data in real-time or near real-time. Depending on the application, this may require output updates every few milliseconds to a few seconds. For instance, a brain computer interface that detects lapses in attention may need to update its output every 100 milliseconds, whereas an offline analysis can afford to average across multiple trials.

This low latency requirement demands algorithmic simplicity and efficiency. As Dr. A explained, certain advanced methods cannot be implemented online. A Butterworth infinite impulse response filter with zero phase distortion, for example, requires both forward and backward processing of the signal and thus access to future data. In real-time, such filters must be replaced with causal, forward-only alternatives, which may introduce slight delays or reduced performance. More generally, any method that requires access to the complete dataset, such as full Fourier transformations or batch model training, must be adapted to support incremental or streaming computation.

Processing Power and Hardware

Offline computations are typically executed on desktop machines with multicore processors, large memory capacity, and sometimes graphical processing units. Cloud computing and high-performance clusters (HPC) are often employed for intensive tasks, such as training deep neural networks on combined EEG and fMRI datasets. In contrast, online systems, particularly those deployed on portable or wearable platforms, face significantly greater hardware constraints. When computations are performed on a laptop or mobile phone during an experiment, it may compete for CPU cycles with other background tasks and be limited by battery life.

Embedded processing on sensor devices introduces even tighter limitations. Many wearable EEG and physiological monitoring devices rely on microcontrollers with restricted memory and processing capabilities. Algorithm design must therefore be tailored to fit these constraints. While offline algorithms may load gigabytes of data into memory, real-time systems typically operate on rolling buffers spanning only a few seconds. Dr. F emphasized that online device computation often compromises other performance aspects, such as increasing battery consumption, heat generation, or device weight. To address these issues, processing code must be optimized using low level programming and efficient libraries and sometimes offloaded to dedicated hardware such as digital signal processors (DSP).

Algorithm Complexity and Fidelity

Another important distinction lies in the complexity and fidelity of algorithms. Offline methods are often optimized for maximum accuracy and detail, whereas online algorithms prioritize efficiency and responsiveness. Dr. E noted that offline analyses produce more complete and higher dimensional representations, while real-time implementations compress the data into lower dimensional forms that are easier to handle. For example, a full EEG spectrogram may be computed offline to characterize cognitive state, while an online implementation might rely on just one or two frequency band metrics, such as beta or alpha power.

This form of data compression involves a tradeoff between richness and utility. It is often necessary to reduce the data to compact, informative indices that enable real-time interpretation. Algorithms must also be explicitly designed to process incoming data in chunks and update outputs incrementally. Dr. D explained that making an algorithm compatible with streaming involves not only algorithmic adjustments but also engineering expertise. Some ML models, especially those designed for batch processing, require modifications to support online learning or inference. Simpler models such as moving window classifiers or incremental learners are often better suited to real-time deployment.

Robustness and Maintainability

Offline analyses are typically conducted in flexible software environments that are easy to update or modify. In contrast, real-time systems deployed in consumer devices may implement algorithms in firmware or hardware to improve performance and energy efficiency. As Dr. B stated, embedding an algorithm in a hardware circuit or firmware can make updates difficult or impossible, which increases the importance of upfront validation and robustness. On the other hand, implementing core computations in hardware can improve reliability and ensure consistent timing, which are critical in closed loop systems.

For example, consider a system that integrates eye tracking and EEG to detect workload and provide real-time feedback. Offline, one might use a complex pipeline involving independent component analysis, wavelet decomposition, and deep convolutional networks to achieve high classification accuracy. However, deploying this pipeline in real-time on a tablet used by a student would be computationally impractical. Instead, the system might use a simplified feature set, such as frontal beta power and blink metrics, with a lightweight classifier such as a linear discriminant function. Although this may result in lower accuracy, the model can operate efficiently and deliver timely feedback.

As Dr. G observed, handling issues such as motion artifacts in real-time is often infeasible with limited processing power. In such cases, system designers may need to accept some loss in accuracy or compensate with well-designed sensors that minimize noise at the source.

Design Implications

Experts highlighted differences in the design philosophy of offline versus online algorithms. Offline analyses are often exploratory, designed to extract maximum insight from data. In contrast, online device algorithms are typically purpose built, streamlined, and tuned for speed and reliability. Dr. F provided the example of a wearable posture sensor designed to alert users when they slouch. Such a device must rely on a simple rule-based algorithm that operates reliably and continuously. Attempting to perform complex diagnostic tasks, such as detecting neurological disorders, on the same hardware would likely compromise its effectiveness. In these cases, a hybrid approach known as edge computing may be employed. Basic processing occurs on the device, while more complex analysis is performed on a remote computer using streamed data or intermediate results.

When developing algorithms for multi-modal cognitive monitoring, it is essential to determine whether the end use is offline analysis, online real-time application, or both. Algorithms intended for real-time deployment should be tested under realistic streaming conditions. Techniques such as load testing, code profiling, and use of compiled programming languages are important to ensure responsiveness and reliability. Often, an algorithm developed offline must be re-engineered for deployment. For example, a MATLAB script (MathWorks Inc, MA) may need to be rewritten in C++ or optimized in Python to achieve the required execution speed. In some cases, a deep learning model may be used offline to identify key features, while a simpler logistic regression model is implemented online for efficient execution.

Ultimately, the distinction between offline and online processing reflects a tradeoff between insight and immediacy. Offline pipelines prioritize comprehensive understanding, while real-time systems aim to deliver actionable results within strict constraints. Both approaches are essential. Offline analysis informs theoretical models and system design, whereas online capability enables practical applications such as neurofeedback, wearable monitoring, and brain computer interaction. As computational hardware continues to advance, the gap between offline and real-time capabilities will narrow. In the meantime, careful algorithm design, optimization, and a clear understanding of system constraints are essential for successful real-time deployment in multi-modal cognitive monitoring.

Conclusion

Multi-sensor neurophysiological recording offers exciting possibilities for understanding cognition in realistic settings, but it comes with substantial technical challenges. In this report, we outlined four key challenges: <u>synchronization</u>, <u>real-time analysis</u>, <u>data fusion</u>, and <u>computational demands</u>, and surveyed current solutions for each. Synchronization across modalities is critical to ensure that disparate signals can be meaningfully compared. Researchers employ hardware triggers and software framework like LSL (Kothe et al., 2024) to achieve timing precision on the order of milliseconds or better. Online analysis requires robust artifact handling and clever validation strategies to maintain data quality and model performance during live experiments, while some progress has been made (e.g., real-time feedback loops and adaptive QC measures), it remains difficult to match offline data cleaning in real-time. Data fusion and integration techniques are unlocking the potential of multi-modal insights. Effectively combining streams (especially in real-time) demands careful feature engineering and still often falls short of its promise, pointing to an important direction for future research. Finally, the computational constraints of real-time, online device processing forces a trade-off between complexity and speed, a reminder that algorithms must be designed with the end-use environment in mind.

Across these domains, the trend is toward greater integration—integration of multiple data sources, integration of real-time loops when analyzing offline (through shadow modes or periodic validations), and integration of algorithm development with hardware capabilities. The experts predict that some problems, like basic multi-sensor time synchronization, will see significant improvement in the next five years, thanks to advancing wireless technologies and standards. Others, like real-time artifact removal in freely moving humans, will likely continue to challenge us. The pursuit of truly mobile, multi-modal cognitive monitoring will continue to drive innovation at the intersection of neuroscience, engineering, and data science. By tackling the challenges outlined here with interdisciplinary approaches, we can study the brain and body in concert in the real world. This will yield richer understanding of cognitive processes and enable the use of wearable sensors to monitor operators' cognitive states in real-time, guide aircraft automation, and predict performance to help prevent potential mishaps.

References

Aghajani, H., Garbey, M, & Omurtag, A. (2017). Measuring mental workload with EEG+fNRIS. *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/fnhum.2017.00359.

Al-Shargie, F., Kiguchi, M., Badruddin, N., Dass, S. C., Hani, A. F., & Tang, T. B. (2016). Mental stress assessment using simultaneous measurement of EEG and fNIRS. *Biomedical Optics Express, 7*(10), 3882–3898.

Borghini, G., Flumeri, G. D., Arico, P., Sciaraffa, N., Bonelli, S., Ragosta, M., Tomasello, P., Drogoul, F., Turhan, U., Acikel, B., Ozan, A., Imbert, J. P., Granger, G., Benhacene, R., & Babiloni, F. (2020). A multi-modal and signals fusion approach for assessing the impact of stressful events on air traffic controllers. *Scientific Reports*, *10*, 8600

Blum, S., Hölle, D., Bleichner, M. G., & Debener, S. (2021). Pocketable labs for everyone: Synchronized multi-sensor data streaming and recording on smartphones with the Lab Streaming Layer. *Sensor, 21*(23), 8135

Cao, T., Sun, J., Guo, H., Tang, J., Wang, Q., & Liu, D. (2021). Design of wearable and portable physiological parameter monitoring system for attentiveness evaluation. *2021 IEEE International Conference on Artificial Intelligence and Computer Applications* (ICAICA), 1207–1213.

Caldwell, J. A. (2005). Mental workload and fatigue in complex performance environments: Introduction to the symposium. *Aviation, Space and Environmental Medicine*, *76*(7 Suppl), B1–B3.

Chen, S., Tang, J., Zhu, L., & Kong, W. (2022). A multi-stage dynamical fusion network for multi-modal emotion recognition. *Cognitive Neurodynamics, 17*(3), 671–680.

Chen, J. C. C. & Ziegler, D. A. (2025). Closed-loop systems and real-time neurofeedback in mindfulness meditation research. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 10*(4), 377–383.

Chu, H., Cao, Y., Jiang, J., Yang, J., Huang, M., Li, Q., Jiang, C., Jiao, X. (2022). Optimized electroencephalogram and functional near-infrared spectroscopy-based mental workload detection method for practical applications. *Biomedical Engineering OnLine, 21*, 9. https://doi.org/10.1186/s12938-022-00980-1

Deligani, R. J., Borgheai, S. B., McLinden, J., & Shahriari, Y. (2021). Multi-modal fusion of EEG-fNIRS: A mutual information-based hybrid classification framework. *Biomedical Optics Express, 12*(3), 1635–1650.

Diaz-Piedra, C., Garcia-Lopez, J., & Catena, A. (2020). Neurofeedback training with a motor imagery task in a combat pilot: A case study. *Applied Psychophysiology Biofeedback, 45*(2), 137–146.

Duffy, M. J., & Feltman, K. A. (2022). *A systematic literature review of operator state detection using physiological measures* (USAARL-TECH-TR--2023-11). U.S. Army Aeromedical Research Laboratory.

Forte, G., Favieri, F., Casagrande, M. (2019). Heart rate variability and cognitive function: A systematic review. *Frontiers in Neuroscience, 13*, 710.

Gao, T., Chen, C., Liang, G., Ran, Y., Huang, Q., Liao, Z., He, B., Liu, T., Tang, X., Chen, H., & Fan, Y. (2025). Feature fusion analysis approach based on synchronous EEG-fNIRS signals: Application in etomidate use disorder individuals. *Biomedical Optics Express*, *16*(2), 382–397.

John, A., Cardiff, B., & John, D. (2024). A review on multisensory data fusion for wearable health monitoring. *arXiv.*

Kothe, C., Shirazi, S. Y., Stenner, T., Medine, D., Boulay, C., Grivich, M. I., Mullen, T., Delorme, A., & Makeig, S. (2024). The Lab Streaming Layer for synchronized multi-modal recording. *bioRxiv*. https://doi.org/10.1101/2024.02.13.580071

Lee, S., Shin, Y., Kumar, A., Kim., M., Lee, H. N. (2019). Dry electrode-based fully isolated EEG/fNIRS hybrid brain monitoring system. *IEEE Transactions on Biomedical Engineering*, 66, 1055–1068.

Li, R., Potter, T., Huang, W., & Zhange, Y. (2017). Enhancing performance of a hybrid EEG-fNIRS system using channel selection and early temporal features. *Frontiers in Human Neuroscience*, *11,* 462.

Li, R., Yang, D., Fang, F., Hong, K, Reiss, A. L., & Zhang, Y. (2022). Concurrent fNIRS and EEG for brain function investigation: A systematic, methodology-focused review. *Sensor*, *22*(15), 5865.

Lim, B., Arik, S. O., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting, 37*(4), 1748–1764.

Lin, C. T., King, J. T., Chuang, C. H., Ding, W., Chuang, W. Y., Liao, L. D., & Wang, Y. K. (2020). Exploring the brain responses to driving fatigue through simultaneous EEG and fNIRS measurements. *International Journal of Neural Systems*, *30*(1), 1950018.

Matthews, G. & Desmond, P. A. (2002). Task-induced fatigue states and simulated driving performance. *The Quarterly Journal of Experimental Psychology Section A, 55*(2), 659–686

Mullen, T., Kothe, C., Chi, Y. M., Ojeda, A., Kerth, T., Makeig, S., Cauwenberghs, G., & Jung, T.-P. (2013). Real-time modeling and 3D visualization of source dynamics and connectivity using wearable EEG. *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).*

NATO RTO-TR-HFM-162. (2012). *Rotary-wing brownout mitigation: Technologies and training*. NATO Research & Technology Organisation.

Niu, Y., Chen, X., Fan, J., Liu, C., Fang, M., Liu, Z., Meng, X., Liu, Y., Lu, L., & Fan, H. (2025). Explainable machine learning model based on EEG, ECG, and clinical features for predicting neurological outcomes in cardiac arrest patient. *Scientific Reports, 15*, 11498.

Nunez, P. L. & Srinivasan, R. (2006). *Electric fields of the brain: The neurophysics of EEG.* Oxford University Press.

Sams, A. S., Yulita, I. N., Wibawa, A. P., Andriani, R., Pramudyo, H., Pramono, W., & Munawwaroh, L. (2023). Multi-modal music emotion recognition in Indonesian songs based on CNN-LSTM, XLNet transformers. *Bulletin of Electrical Engineering and Informatics, 12*, 355–364.

Saxby, D. J., Matthews, G., Warm, J. S., & Hitchcock, E. M. (2013). Development of a revised NASA-TLX (Task Load Index): Results from a field test of cognitive and perceptual workload. In *Advances in Human Aspects of Aviation* (pp. 495–504). Springer.

Shin, J., von Lühmann, A., Kim, D.-W., Mehnert, J., Hwang, H.-J., & Müller, K.-R. (2018). Simultaneous acquisition of EEG and NIRS during cognitive tasks for an open access dataset. *Scientific Data*, 5

Smith, K., Endsley, T., & Clark, T. (2023). Physiological correlates of objective situation awareness measurements. *2023 IEEE Aerospace Conference,* 1–6.

Su, W. C., Dashtestani, H., Miguel, H. O., Condy, E., Buckley, A., Park, S., Perreault, J. B., Nguyen, T., Zeytinoglu, S., Millerhagen, J., Fox, N., & Gandjbakhche, A. (2023). Simultaneous multi-modal fNIRS-EEG recordings reveal new insights in neural activity during motor execution, observation, and imagery. *Scientific Report, 13*(1).

Sun, X., Dai, C., Wu, X., Han, T., Li, Q., Lu, X., & Yuan, H. (2021). Current implications of EEG and fNIRS as functional neuroimaging techniques for motor recovery after stroke. *Medical Review, 4*(6), 492–509.

Uchitel, J., Vidal-Rosas, E. E., Cooper, R. J., & Zhao, H. (2021). Wearable, integrated EEG-fNIRS technologies: A review. *Sensor, 21*(18), 6106.

Vogl, J., Delgado-Howard, C., Plummer, H., McAtee, A., Hayes, A., Aura, C., St. Onge, P. (2023). *A literature review of applied cognitive workload assessment in the aviation domain* (USAARL-TECH-TR--2023-04). U.S. Army Aeromedical Research Laboratory.

Xefteris V., Dominguez, M., Grivolla, J., Tsanousa, A., Zaffanela F., Monego, M., Symeonidis, S., Diplaris, S., Wanner, L., Vrochidis, S., & Kompatsiaris, I. (2023). A multi-modal late fusion framework for physiological sensor and audio-signal-based stress detection: An experimental study and public dataset. *Electronics, 12*(23), 4871.

Xiao, R., Ding, C., & Hu, X. (2022). Time synchronization of multi-modal physiological signals through alignment of common signal types and its technical considerations in digital health. *Journal of Imaging*, *8*(5), 120.

Zou, C, Deng, Z., He, B., Yan, M., We., J., & Zhu, Z. (2024). Emotion classification with multi-modal physiological signals using multi-attention-based neural network. *Cognitive Computation and Systems*. https://doi.org/10.1049/ccs2.12107

Appendix A. Acronyms and Abbreviations

ADC	Analog-to-Digital Converter
AI	Artificial Intelligence
ANS	Autonomic Nervous System
API	Application Programming Interface
ASR	Artifact Subspace Reconstruction
BCI	Brain Computer Interface
BLE	Bluetooth Low Energy
CCA	Canonical Correlation Analysis
CPU	Central Processing Unit
DSP	Digital Signal Processor
ECG	Electrocardiogram
EDA	Electrodermal Activity
EEG	Electroencephalogram
EOG	Electrooculography
fNIRS	functional Near-Infrared Spectroscopy
FVL	Future Vertical Lift
GLM	Generalized Linear Model
GPS	Global Positioning System
GPU	Graphic Processing Unit
HRV	Heart Rate Variability
HPC	High-Performance Clusters
IEEE	Institute of Electrical and Electronics Engineers
ICA	Independent Component Analysis
IIR	Infinite Impulse Response
IMU	Inertial Measurement Unit
I/O	Input/Output
LE	Low Energy
LSL	Lab Streaming Layer
MEG	Magnetoencephalography
ML	Machine Learning
MoBI	Mobile Brain/Body Imaging
NTP	Network Time Protocol
OSM	Operator State Monitoring
POV	Point of View
PPG	Photoplethysmogram
PTP	Precision Time Protocol
QC	Quality Control
RSP	Respiration
SCG	Seismocardiogram
TFT	Temporal Fusion Transformer
TTL	Transistor-Transistor Logic
USAARL	U.S. Army Aeromedical Research Laboratory
UWB	Ultra-Wideband
VR	Virtual Reality
XDF	eXtensible Data Format

Appendix B. Experts' Responses

1) What level of inter-stream temporal error is still acceptable for your analyses?

Dr. A:

This depends on the streams' sampling rate. For most paradigms target ≤ 10 ms jitter and ≤ 5 ms drift per minute between streams that index fast neurophysiology (EEG, fNIRS Hb O phase locked tasks, pupillometry, EMG).

Cognitive state metrics (e.g., workload indices) tolerate slightly looser alignment ($\pm$ 20 ms) because we generally aggregate across hundreds of milliseconds.

For event locked evoked responses, especially when fusing EEG with stimulus triggers or eye blinks, I insist on sub-millisecond trigger precision even if the continuous streams drift by a few milliseconds over time.

Dr. B:

It really depends on the dependent variable I am interested in (EEG response to fixation onset: lower temporal error required compared to the correlation between facial expression and HRV in a fatigue experiment).

Dr. C:

less than 10 ms

Dr. D:

Generally, if there is a constant delay, we aim at fixing this offline by correcting for it.

For jitter, I would assume not more than 5 to 10 ms (absolute error)

Dr. E:

I am mostly working with EEG, MEG and intracranial EEG in human participants to study cognitive processes. For me I would say this is 5 ms.

Dr. F:

It strongly depends on the nature of the data being collected but, in general, a temporal asynchrony in the order of hundreds of milliseconds is acceptable for fNIRS/DOT data, and in the order of a few milliseconds may be acceptable for EEG data. Other physiological data streams (EKG, respiration, GSR) also have their own features, but I'd say that tens to one hundred milliseconds could be deemed acceptable (where less is better, obviously).

Dr. G:

~1 millisecond

Dr. H:

For some data collection, an error under 20-30 ms is acceptable; however, in many of our studies we're looking at smaller event related potentials based on gaze-contingent stimuli. For those, error greater than 10 ms could cause sufficient variability to drown out smaller components of interest.

2) If you have unlimited resources, what would be your "gold-standard" synchronization setup look like? Why isn't it practical today?

Dr. A:

Ideal world:

- Precision time protocol (PTP/IEEE 1588) grand master clock distributed over fiber to every sensor hub, with hardware level timestamping at ADC sampling, plus shared PPS (pulse per second) derived from GPS to eliminate drift.
- Redundant trigger lines (TTL) and optical isolators for fail safe.
- Central ledger that logs clock skew estimates in real-time so software can dynamically resample.

Why it's impractical today:

- Wearable sensors run on BLE and/or limited chipsets that lack deterministic timestamping.
- Fiber or PPS cabling negates mobility and ecological validity.
- Sub µs hardware clocks inflate power draw and cost; most head worn devices can spare only a few milliwatt (mW) for timing.

To account for these limitations and improve on the existing solutions, actually, we have developed a hardware solution in my lab. I have advised three different master theses over the years (listed below) as we iterated it.

MS theses that I advised for the hardware solution that we call NeuroHub and its evolution:

Grzeczkowski, N. V. (2014). *NeuroHub: Portable and scalable time synchronization instrument for brain-computer interface and functional neuroimaging research* [Master's Thesis, Drexel University].

Thomas, N. (2017). *NeuroHub networking integration: Time synchronization device for multi-modal brain imaging and hyperscanning research* [Master's Thesis, Drexel University].

Dai, A. G. (2021). *NeuroHub fog: Wireless network time synchronization device for multi-modal brain imaging and hyperscanning research* [Master's Thesis, Drexel University].

Dr. B:

The gold standard synchronization setup is very achievable by means of widely supported Lab Streaming Layer (LSL) framework and attainable networking equipment. I have been working with LSL since it was published about 6–7 years ago now and my gold standard would be every experiment apparatus out-of-the-box supporting it.

Dr. C:

Stream sync should be done directly on the sensing hardware to avoid network latencies. Timestamp alignment between streams will always have failure modes.

Dr. D:

All tethered to one PC—not practical as it won't allow naturalistic movements of participants in most of the hardware settings we use and/or the devices that allow mobile recordings cannot be tethered

Dr. E:

Parallel port type of cable with TTL level signals. Well, that is the old-fashioned way that works robustly in lab settings... Perhaps with unlimited resources I would want to use parallel optical fibers in the same way as TTL signals over a multi-core cable, or possibly serial communication over a single fiber with parallel TTL interfaces on both ends. The length of the fiber would not be a problem and there is no electromagnetic interference. That makes it compatible with EEG, MEG, fMRI, but also in harsh and less controlled environments like a clinical setting.

I can think of a few different reasons why it is not practical:

The parallel cable is not practical because digital communication between devices (and chips) has switched to serial. Serial-over-USB is the standard and Arduino-like devices are needed to get triggers from button-boxes and to amplifiers.

Regarding optical fibers, lab support staff can manage a soldering iron and an oscilloscope, but optical interconnections are harder to implement, test, and maintain in an environment where the setup needs to be flexible (like a research lab, where very regularly new equipment needs to be incorporated into the setup).

Parallel TTL signal is also not practical as it limits the representation of information that can be conveyed to numbers, often between 1-255 or 1-(2^16-1). Coding of more complex information (for example, where on the screen did what stimulus appear?) requires numeric sequences of trigger codes, whereas more flexible representation (as allowed in the BIDS events.tsv or with HED tags) is more practical.

Labs are getting more complex and wired setups are hard to implement and maintain. Wireless allows for more flexible lab setups. LSL over Wifi allows for that, but sometimes we also use ZeroMQ (Zero Message Queue).

Dr. F:

A wireless setup with optimal performance would be ideal in many situations. In essence, an LSL interface without the complexities of wired networking and protocols.

Dr. G:

Wireless mechanisms, using a standard such as Bluetooth low energy (BLE) to allow broad use of consumer electronics gadgetry. BLE is not ideally set up for sub-millisecond wireless time synchronization, unless separate local clocks are built into the devices, which requires additional power burden.

Dr. H:

Currently we're building the infrastructure to record from IMUs on multiple rifles and stream data from multiple HoloLenses. The "gold standard" that we're aiming for is to be able to record from larger scale units. Currently this is difficult to manage due to bandwidth issues.

3) Which artifacts are still hardest to suppress in real-time without post-hoc processing?

Dr. A:

Motion coupled hemodynamic fluctuations in fNIRS (especially scalp blood flow impact).

EMG contamination of high gamma EEG during speech or chewing.

Respiration linked baseline drift in wearable ECG when subjects are talking or moving.

Real-time regression filters help but truly separating physiological covariates without post-hoc ICA/Generalized linear model (GLM) remains tough.

Dr. B:

Even with post-hoc processing particularly, distinguishing the dependent variables we are interested in from the noise floor can still be problematic. With push for ecological validity and more realistic experiment settings ever on the rise in our field of neuroergonomics, the noise floor has been on a trend of increase alongside. Environmental noise is still therefore, the biggest challenge. This ranges from easier fixes, such as continuing to use gel ExG electrodes instead of dry to reduce movement artefacts or putting isolation against sunlight in eye tracking or fNIRS studies to random radio frequency jammer of some bureaucrat driving nearby and disrupting experiments.

Dr. C:

It depends a lot on the modality. EEG is not affected by the same artifacts as EMG for example. It also depends a lot on what you do with the data. If the data gets piped to a machine learning maybe you don't need to clean anything.

Dr. D:

Mechanical artifacts due to cable sway or the electrode skin interface (impedance changes).

Dr. E:

Movement artifacts where the transducer loses touch with the skin.

Dr. F:

Movement artifacts in neuroimaging are not so difficult to detect in real-time, but correcting them on the fly is certainly a higher-level challenge.

Dr. G:

Episodic artifacts from motions.

Dr. H:

Extracting head movement artifacts (both noise from wire movement as well as muscle activity) is difficult to parse in real-time.

4) Do you trust real-time quality metrics? Why?

Dr. A:

I treat them as early indicators, not final results. Signal quality indices based on impedance, SNR, or kurtosis are reliable for gross failures (open lead, saturated diode) but poor at detecting subtle non stationarities. We still log full session data and run offline diagnostics before drawing conclusions.

Dr. B:

Depends on the dependent variable again, in that, if what is perceivably real-time can be tractably analyzed and output during data flow – great. For instance, in EEG, environmental factors can be strong enough to outline entire wavebands, in which case having a quality metric that would distinguish environmental noise, within device noises (I have some first-hand experiences with bad circuit designs causing imperfect data recordings), and participant related noise – then that would be a real-time quality metric worth a look. And this is easier to attain with due process than realized. But an EEG device reporting on impedance? Not as much as I trust my trained eye.

Dr. C:

With the setup I use, yes, I do.

Dr. D:

We have not been working with real-time quality metrics yet and I would be doubtful that they would be reliable and robust for different movement scenarios.

Dr. E:

Real-time QC is useful if it leads to real-time actions of the experimenter in case the quality is not as good as it should be. For example, the experimenter re-attaching an electrode. Real-time loses its value if it is open-loop and if there is no feedback to the measurement.

Real-time QC can also be useful if more information is available during the measurement than what is encoded and stored on disk. This may happen for example with eye trackers, where the real-time QC can use the camera image, whereas the signal stored on disk has been condensed into the x and y gaze position and blinks.

If the real-time signal on which the QC is computed is the same as the signal that is available offline, then offline always has the advantage that there are no computational limits (as in real-time) and that there is flexibility in the choice of the algorithm or parameters.

Dr. F:

Several metrics for real-time assessment of data quality are available, and in most cases are reliable. Obviously, a post-hoc analysis of data quality is also necessary to put in the context of the experiment, and generate a proper strategy for data pruning or filtering.

Dr. G:

If real-time means updates on tens of seconds timescales to allow for averaging and rejection of outliers, then yes. If real-time means second to second, or sub-second, measurements, then no.

Dr. H:

It depends. On homemade data collection platforms it can be easier to classify bad samples on the fly. However, depending on the depths of the black box in the vendor's application programming interface (API) there could be interpolation methods that mask bad samples and are not able to be flagged/rejected until post-hoc processing.

5) What is the biggest difference between an algorithm that performs well offline and one that survives on-device deployment?

Dr. A:

Offline processing affords different types of filters (such as infinite impulse response (IIR), Butterworth with zero-phase correction, that is done in both forward and backward passes for zero phase correction and requires all data at hand) and cannot be done online in real-time.

Dr. B:

Under the assumption that on-device deployment of an algorithm has been done on a circuit level, then circuit designs' robustness vs. inflexibility and tangibility is a double-edged sword; yes, robustness can be far greater in controlled and well tested circuitry but the device circuit needing anything beyond a firmware update? That would render it unusable, therefore unfeasible. Assuming a software deployment of on-device processing comes with a plethora of tractability, hardware design, cost, and maintenance problems of its own.

Dr. C:

If the target is one device usage with streaming data you need to make sure that any algorithm you use is stream compatible and implemented as such. Then putting this on device should not lead to any blocker beyond money and skills.

Dr. D:

Higher flexibility for most offline algorithms

Dr. E:

Offline algorithms can provide the researcher with more complete but also more complex representations of the signal features of interest. An on-device algorithm usually serves to "compress "the high-dimensional signal into something that is as lo-dimensional as possible, as that is the easiest to interpret and work with.

Dr. F:

It strongly depends on what the goal of the algorithm is. For instance, an on-board quality checker with immediate feedback to the user is impactful, and it therefore it must be both reliable and computationally efficient. However, if the algorithm's objective is data processing/analysis leading to interpretable results, an on-device implementation is not only difficult to implement but also may compromise other useful device features like battery duration, computational demand, weight, and other ergonomical and practical aspects.

Dr. G:

Algorithms that can handle motion artifacts are needed for on-device deployment, often difficult given the power and compute constraints.

Dr. H:

Ability to handle out of distribution or rare events.

6) How do you validate a closed-loop ML model that adapts on the fly?

Dr. A:

Shadow mode logging: Run the adaptive controller in parallel while the actual stimulus is driven by a trusted baseline; compare decisions post-hoc.

Replay benchmarking: Feed the exact recorded streams back through multiple frozen checkpoints of the model to audit divergence.

Interleaved lock in trials: Insert periodic calibration epochs where ground truth labels are known (e.g., visual oddball) so you can quantify drift without interrupting the study.

Dr. B:

I answer under the assumption that this is not a fully black-box approach more so followed by deep learning aficionados, and instead a well-designed complex feedback mechanism of a multi-layered ML model. We can go two methods here; a theoretical approach where the mathematics behind the ML is proven, the implementation is sound and valid, and the overall realization can be tested within tractable circumstances, and a testing data set to also provide an empirical perspective to testing for controlled as well as uncontrolled input/output (I/O).

Dr. C:

Via behavioral performance, e.g., how fast the person can do the task.

Dr. D:

Not yet worked with such an approach but I would try to come up with a standardized test-battery that includes different standardized movements.

Dr. E:

It is not clear to me here what the closed-loop aspect of the ML model is. There are many loops that can be closed with physiological signals, but whether closing them is meaningful depends a lot on the real-world application or experimental setup.

I would distinguish between validity and reliability. I would assess reliability by repeatedly evaluating the ML model on different versions of the signal with varying signal quality. That does not have to be done online. I would test validity using some sort of experimental design where the outcome of the ML model would be specifically studied under known experimental perturbations (e.g., conditions that the ML model is meant to detect).

Dr. F:

I don't do much ML in my work, so I can offer just a general and perhaps obvious perspective. Validation should involve observations that are distinctively dissimilar from the training set, so to understand the generalizability of the model and, as a subsequent step, to develop a refinement/adaptation of such model for better performance.

Dr. G:

Probably start with validation in a controlled setting such as a hospital, then progressing to home use – ultimately to the field.

Dr. H:

No response

7) How do you handle missing packets or corrupt segments in a long recording when the study cannot stop?

Dr. A:

Reject/eliminate that part of the task condition(s) from analysis.

Dr. B:

Depends on the dependent variable, the duration of data corruption and/or the amount of missing packets. If it is too valuable data, gone for too long, then I would go as far to drop portions of experiment blocks altogether if I absolutely cannot stop and/or restart the study. However, that is bad design to begin with and I hope to never do one. If it is so much so that we end up with too little a dataset, then there have been novel methods within the past few years for generative-adversarial-networks (GANs) to more traditional ML approaches, and even signal processing theory to visit to recreate, repair, or interpolate the data if the data type fits this frame of possibilities.

Dr. C:

If the data is lost, it's lost.

Dr. D:

These are usually removed offline from analyses in most our approaches as we have no ground truth data for different movements that would allow for some kind of interpolation.

Dr. E:

I would try to identify that they are missing (for example by looking at jumps in a clock or counter channel) and not interpret the data that spans over such an interruption, i.e., the algorithm would skip that section of data. I would not try to interpolate the data.

Dr. F:

In task-based experiments, we typically remove/trim entire data segments that may lead to false positive results, i.e., an entire block, or trial.

Dr. G:

Usually these time segments are simply not analyzed.

Dr. H:

Currently we don't have an agreed upon method on how to handle missing packets in such a scenario. Typically, this is just flagged and handled in post-processing.

8) What is the best method you've seen for ground-truthing multi-modal physiological streams in the wild?

Dr. A:

A hybrid of synchronized head mounted video + inertial measurement + stamped experimenter annotations. The video provides context; IMU gives frame accurate movement onset; annotations mark task events. When all three agree, you have a defensible ground truth.

Dr. B:

Divide and conquer has been one. Where all dependencies of a very complex interaction setting can be broken down to their fundamental components (as far as our perception of the interaction setting allows of course) and tested individually. Done with diligence and controlling for unknown interaction effects as best as possible (i.e., not using components that are too intertwined) then this works well. For instance, using an experiment battery to investigate how outcomes would be affected by the interaction environment and more to see whether the experiment apparatus can still perform functionally enough.

Dr. C:

You could run quality checks online, but this has a computational cost. It also depends a lot on what stream you consider.

Dr. D:

I have not seen any ground-truthing approach for multi-modal but only single tests for single modalities

Dr. E:

Comparing it to golden-standard physiological streams in well-controlled lab environments.

Dr. F:

LSL offers good implementation for multi-modal streams but, testing wise, custom methods are often employed by individual researchers/labs.

Dr. G:

Measurements with multiple modalities or devices, e.g., HR from ECG, SCG, and PPG, to establish redundancy and internal consistency.

Dr. H:

Collecting point of view (POV) video data has been helpful for verify data collection in eye tracking and head tracking.

9) What is the most overlooked privacy, security risk when streaming raw physiological data wirelessly?

Dr. A:

Physiological "side channel" identity leakage. Even if names are stripped, raw ECG/EEG/fNIRS spectra contain quasi unique biometric signatures. Streaming them over lightly encrypted BLE (common in consumer wearables) enables re-identification attacks.

Dr. B:

The use of Bluetooth low energy (LE) is often done without much care due to how little range it already, perceivably allows – I mean, it is a standard for fully unencrypted data transfer without as much as a client-server handshake over the air that capturing it is a thumb sized USB device away. Making sense of an unknown data structure is still madness though, and with quality critical data – it is easy to understand why there is not much care given to it. But other than that, 802.11 can be tractably bullet proof for the use we have for it while regular Bluetooth is strong enough.

Dr. C:

It's pretty obvious that you record the packets on the Wifi network and if the data is not encrypted you can collect the data without authorization.

Dr. D:

Not sure, potentially other devices recording the data that might further allow personal information to be matched with the physiological data (e.g., photos).

Dr. E:

That being near the person being recorded (i.e., in wireless range) might disclose information about the stream to a bad actor. That the stream is transmitted through an external device (e.g., first Bluetooth to phone, then from phone via Wifi to the cloud) and that the in-between device can be the weak link. That data gets stored on an external device (like a phone, computer, or cloud). The number of bad actors able to connect to the cloud (or to the in-between device) is infinitely larger than the number of bad actors that can be in physical proximity.

Dr. F:

Associating eavesdropped data to participants is certainly a concern, although decoding and interpreting the data stream in a way that violates privacy remains a challenge for bad actors.

Dr. G:

Don't know how to answer this one.

Dr. H:

It's not necessarily overlooked but obviously streaming live data wirelessly is vulnerable. As with any data collection, it is important to remember that even if data does not contain PII, it can be reidentified if enough data is collected and the sampling method for participants is restricted.

10) Looking five years ahead, which current research question will feel "resolved" and which challenge will still be with us?

Dr. A:

"Probably solved"

Low jitter multi sensor time sync for wearables; next gen BLE/ULPW UWB chipsets already demo sub microsecond network timing.

"Still haunting us"

Real-time motion artifact rejection for fNIRS/EEG in fully free-living conditions. Human movement diversity outpaces current online models, and power efficient optical flow or EMG aware correction remains an open challenge.

Dr. B:

No response

Dr. C:

Provided enough physiological data are collected in a consistent setting (same hardware, same setup) we'll be able to have models that generalize to any person without personalized models.

The challenge will be to address hardware variabilities and heterogeneous datasets.

Dr. D:

Resolved, hardware development for improved and synchronized multi-modal recordings.

Open issues: Data fusion approaches allowing for “real” multi-modal analyses. "real" in the sense that the data is fused, and different modalities inform the analyses of other modalities. We often record multi-modal but then analyze EEG alone, ECG alone etc.

Dr. E:

Miniaturization of electronics will allow more embedded sensing devices to be built into other things, hence allowing for more multisensory data capture.

Researchers will still not have a better training in or understanding of the signals that they are working with. Capturing the experimental conditions in which data was recorded (i.e., annotations and metadata) will still be limited.

Dr. F:

Noise and movement artifacts will not disappear, but I believe that great progress could be made in correcting or trimming unusable data very reliably, especially using multi-modal data that support one another, i.e., wearable IMU data and neuroimaging data.

Dr. G:

Conventional vital signs may be resolved. Motion artifact cancellation and accurate biochemical measurements will continue to be difficult.

Dr. H:

Team data collection should be solved within the next 5 years (e.g., multisensory data streams simultaneously collected from multiple members of the squad).

U.S. ARMY AEROMEDICAL RESEARCH LABORATORY

FORT RUCKER, ALABAMA

Optimizing

HUMAN PROTECTION AND PERFORMANCE

since 1962

All of USAARL's science and technical informational documents are available for download from the Defense Technical Information Center.

https://discover.dtic.mil/results/?q=USAARL

U.S. ARMY

T2COM

MRDC