System Description Paper: SSVEP-based Brain-Computer Interface for MTC AI Competition

Team Imitation

2025

Abstract

This paper presents a comprehensive technical documentation of a high-performance Steady-State Visual Evoked Potential (SSVEP) based Brain-Computer Interface (BCI) system developed for the MTC AI Competition. Our approach emphasizes advanced signal processing and feature engineering over purely deep learning solutions, achieving a Leave-One-Subject-Out (LOSO) F1-score progression from 0.67 to 0.715, with expected improvements beyond this threshold through strategic outlier exclusion. The system integrates a diverse suite of features including Filter Bank Canonical Correlation Analysis (FBCCA), Task-Related Component Analysis (TRCA), Phase-Locked Value (PLV), harmonic analysis, Signal-to-Noise Ratio (SNR), Power Spectral Density (PSD), and template correlation features. These are combined through an ensemble learning approach utilizing XGBoost, Support Vector Classifier (SVC), and Linear Discriminant Analysis (LDA). Our chronological development narrative reveals critical insights about the limitations of deep learning in noisy EEG environments and the paramount importance of deterministic signal processing for achieving robust generalization across subjects. The work demonstrates that for SSVEP classification, the challenge is not merely feature extraction but rather finding the "needle of signal in a haystack of noise," requiring sophisticated domain-specific approaches rather than generic machine learning solutions.

1 Introduction & Problem Statement

1.1 Background & Motivation

Brain-Computer Interfaces (BCIs) offer direct communication between the brain and external devices. Steady-State Visual Evoked Potentials (SSVEPs) are a prominent BCI paradigm due to their high signal-to-noise ratio and ease of implementation. This paper details the development of a robust SSVEP-based BCI system for the MTC AI Competition, aiming for superior classification accuracy and generalizability.

1.2 Problem Definition

The core problem is accurate and rapid classification of user intent from SSVEP responses, amidst significant challenges:

- **Low Signal-to-Noise Ratio (SNR) in EEG:** EEG signals are inherently noisy, making SSVEP extraction a "needle in a haystack" problem, requiring sophisticated signal processing.
- **Inter-subject Variability:** SSVEP responses vary significantly across individuals, necessitating robust generalization.
- **Real-time Processing Requirements:** Practical BCI applications demand near real-time data processing.

- **Robustness to Artifacts:** The system must be resilient to various EEG artifacts.

The following figure shows the EEG Signal Processing challenges ranked by complexity.
The following figure shows the EEG Signal Processing challenges ranked by complexity.

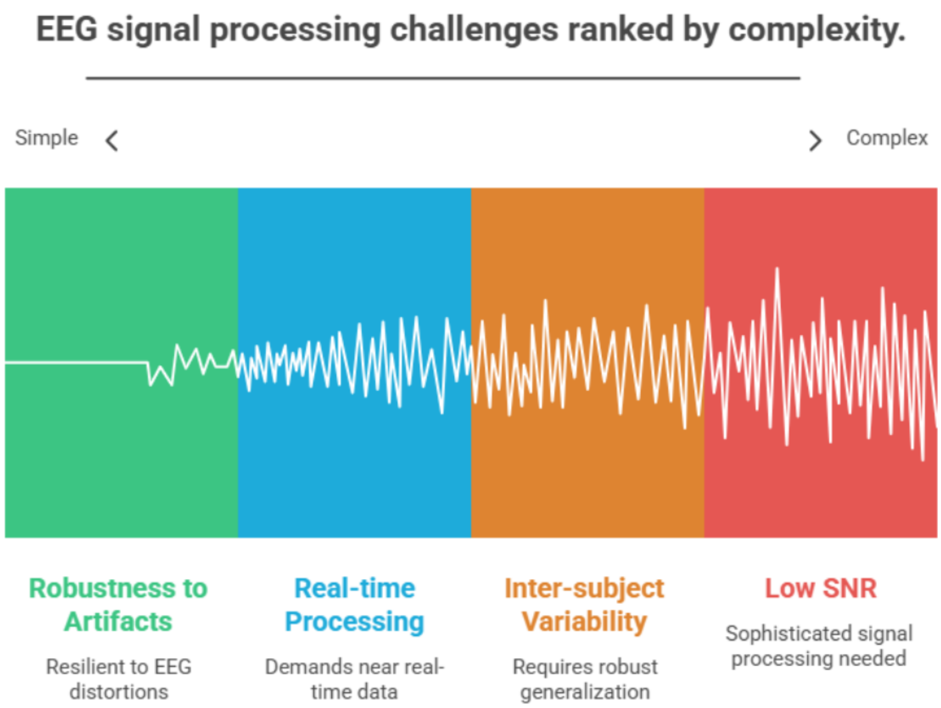

Figure 1: EEG signal processing challenges ranked by complexity.

1.3 Objectives of the Study

Primary objectives include developing a highly accurate and robust SSVEP classification algorithm, minimizing inter-subject variability, achieving high competitive performance, and providing comprehensive documentation of the development process.

2 Related Work Survey & Challenges

2.1 Literature Review Overview

SSVEP BCI research has evolved from basic frequency-domain analysis to sophisticated methods like Canonical Correlation Analysis (CCA) [1], Filter Bank Canonical Correlation Analysis (FBCCA) [2], and Task-Related Component Analysis (TRCA) [3]. These techniques enhance SSVEP components and suppress noise, improving classification accuracy.

2.2 Strengths and Weaknesses of Prior Work

Table 1: Comparison of different approaches in SSVEP BCI research

Topic	Strengths	Weaknesses
Deep Learning Methods (e.g., CNNs, LSTMs)	The primary advantage is their ability to learn hierarchical features directly from raw data, potentially capturing subtle patterns that might be missed by handcrafted features. They can also be highly flexible and adaptable to different data distributions with sufficient training data.	Deep learning models often require vast amounts of labeled data for effective training, which can be a significant limitation in BCI research where data collection is expensive and time-consuming. They can be prone to overfitting, especially with noisy EEG data. Crucially, as observed in our experiments, deep learning models can struggle to differentiate true signal from noise in highly contaminated EEG, often learning from the background noise rather than the actual SSVEP signal.
Traditional Methods (e.g., CCA, FBCCA, TRCA)	These methods are generally robust, interpretable, and computationally efficient. They leverage established principles of signal processing to effectively extract relevant SSVEP features. Their performance is often competitive with, and sometimes superior to, more complex deep learning models, especially with limited training data.	They require domain expertise for feature engineering and parameter tuning. Their performance can be sensitive to individual differences and variations in experimental conditions if not properly adapted.
MOABB	A key strength was identifying an external resource that provided access to SSVEP datasets and multiple classification pipelines, demonstrating high accuracy and offering guidance on effective model design. Also, MOABB used within subject evaluation	A notable limitation was that the reference data were exceptionally clean, unlike our noisier dataset. This raised concerns about the direct applicability of those methods to our setting, highlighting the challenge of generalizing from ideal conditions to real-world scenarios.

2.3 Gaps in the Literature

Public EEG datasets are often so clean that the primary limitation lies in the model design rather than the data itself. This creates a gap when transitioning to real-world applications, where noise and variability are more prominent. Key challenges include ensuring robustness to real-world noise, achieving true inter-subject and inter-session generalization, and enabling deep learning models to perform reliably under less controlled conditions. Additionally, many existing resources lack detailed, step-by-step development narratives, making reproducibility and adaptation more difficult.

3 Methodology

3.1 Research/Design Approach

Our SSVEP BCI system prioritizes robust signal processing and feature engineering over black-box deep learning. The methodology follows a multi-stage pipeline:

1. **Data Preprocessing and Outlier Exclusion:** Cleaning raw EEG data and removing poor-quality trials/subjects.

2. **Advanced Feature Engineering:** Integrating sophisticated techniques like CCA, TRCA, PLV, harmonic analysis, and SNR estimation to extract relevant, noise robust SSVEP features.

3. **Ensemble Classification:** Combining multiple diverse classifiers for enhanced accuracy and robustness.

4. **Subject-Specific Adaptation:** Incorporating mechanisms for subject-specific template generation and spatial filtering to adapt to individual physiological characteristics.

3.2 Tools, Frameworks, and Technologies Used

The development and implementation of our SSVEP BCI system leveraged a combination of standard scientific computing libraries and machine learning frameworks in Python. The primary tools and technologies include:

- **Python:** The primary programming language for all development.

- **NumPy:** Essential for numerical operations, especially array manipulation and mathematical computations on EEG data.

- **Pandas:** Used for efficient data loading, manipulation, and management of tabular data, including trial metadata and outlier lists.

- **SciPy:** A fundamental library for scientific computing, providing critical modules for signal processing and linear algebra.

- **Scikit-learn (sklearn):** A comprehensive machine learning library utilized for various tasks, including:
  - Preprocessing: StandardScaler for feature scaling and LabelEncoder for transforming categorical labels.
  - Dimensionality Reduction/Feature Extraction: CCA (Canonical Correlation Analysis) and LinearDiscriminantAnalysis (LDA).
  - Classification: SVC (Support Vector Classifier) and VotingClassifier for ensemble learning.

- **XGBoost:** A highly optimized gradient boosting library (xgb.XGBClassifier) known for its speed and performance in classification tasks.

- **Pickle:** Used for serializing and deserializing Python objects, enabling the saving and loading of the trained model and its associated assets (scalers, templates, filters).

3.3 Data Collection and Preprocessing

The dataset comprises 8-channel EEG recordings at 250 Hz. Preprocessing involves:

- **Data Loading and Trial Segmentation:** Extracting 4-second SSVEP segments after a 3-second skip to focus on steady-state responses and stimulation of the brain signals.
- **Common Average Reference (CAR) Preprocessing:** Applying CAR to reduce common noise sources across channels.
- **Outlier Exclusion:** Systematically removing outlier trials/subjects (e.g., subject 27) based on outliers_list.csv to improve model generalization.

3.4 Model Architecture / Algorithm Details

Our model combines a sophisticated feature engineering pipeline with an ensemble classifier.

3.4.1 Feature Extraction Pipeline & How They Complement Each Other

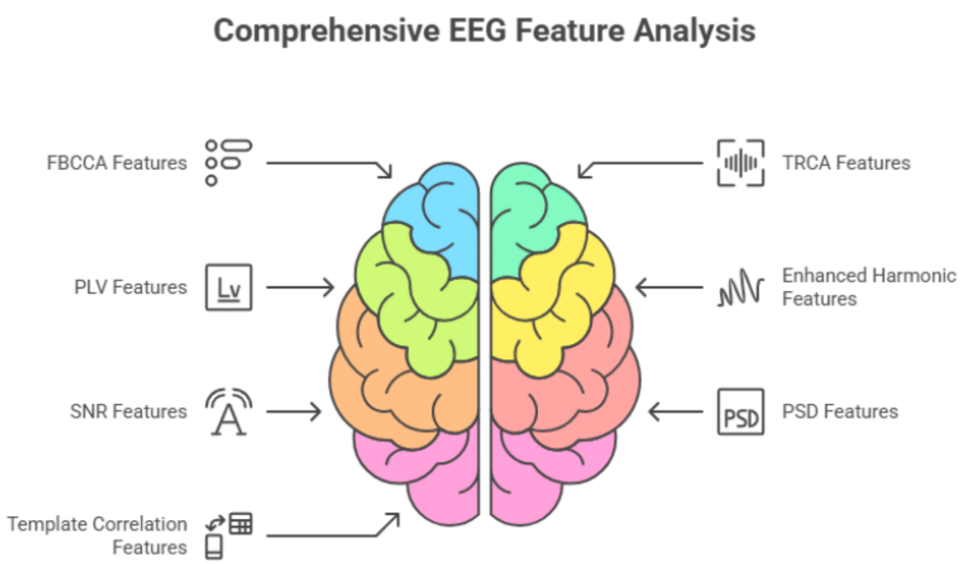

Figure 2: Comprehensive EEG Feature Analysis showing different feature extraction methods

1. **FBCCA Features**
   - Capture correlation between EEG subbands and reference signals at fundamental and harmonic frequencies.
   - *Strength:* Dominant frequency detection and harmonic power.
2. **TRCA Features**
   - Apply spatial filters to maximize reproducible components across trials.
   - *Complement:* TRCA isolates reliable brain responses that FBCCA may dilute by aggregating frequency bands.
3. **PLV Features**
   - Measure phase consistency across channels or between EEG and stimulus.

- *Complement:* Adds dynamic phase-coupling info—FBCCA and TRCA capture amplitude/spatial structure, PLV captures timing coherence.

4. **Enhanced Harmonic Features**
   - Quantify power in fundamental and harmonic bands beyond what FBCCA provides.
   - *Complement:* Reinforces FBCCA's frequency detection with explicit harmonic power statistics.

5. **SNR Features**
   - Estimate how prominent the SSVEP frequency is against neighboring noise.
   - *Complement:* While FBCCA/TRCA focus on template matching, SNR adds a confidence metric based on signal clarity.

6. **PSD Features**
   - Capture the spectral envelope around target frequencies.
   - *Complement:* Provides a fine-grained frequency-domain overview that supports FBCCA's coarse correlation.

7. **Template Correlation Features**
   - Compare trials to subject-specific/global templates using correlation and phase metrics.
   - *Complement:* Adds personalized consistency checks that help generalize across subjects when combined with ensemble spatial and frequency measures.

3.4.2 Ensemble Classification Model

We employed a VotingClassifier (soft voting) composed of three complementary models, each selected for its distinct strengths. We also performed a grid search to find optimal weights, resulting in the final configuration: XGBoost (0.5), SVC (0.3), and LDA (0.2). This setup ensures balanced, robust performance as shown in the following figure:

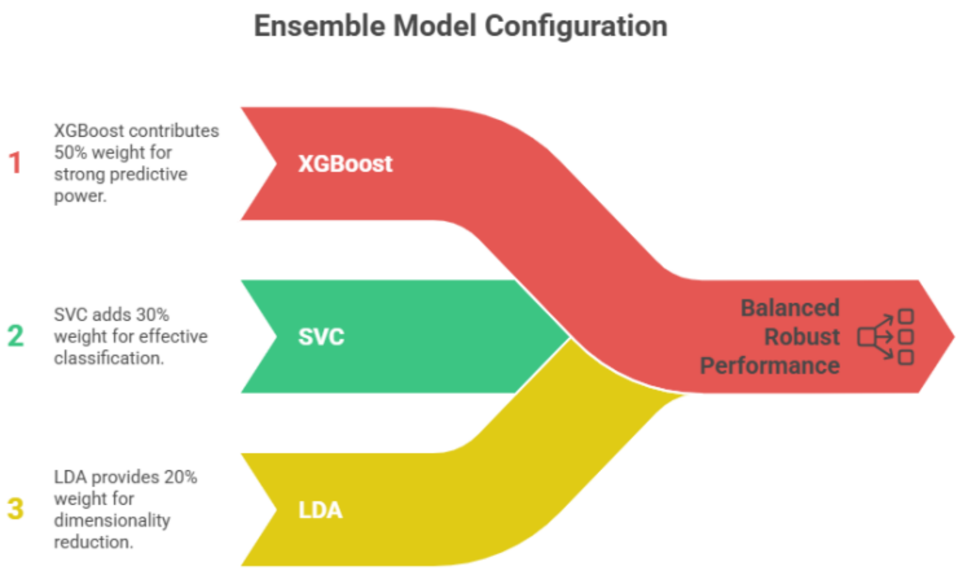

Figure 3: Ensemble Model Configuration showing the three classifiers and their weights

1. **XGBoost Classifier (weight = 0.5)**
    - Excels in capturing complex, nonlinear interactions through gradient-boosted decision trees
    - Achieves the highest average accuracy, although it may struggle with certain subjects due to overfitting on noisy data or outliers.
    - Holds the most influence in the ensemble due to its overall effectiveness.
2. **Support Vector Classifier (SVC, weight = 0.3)**
    - Robust at finding non-linear decision boundaries via kernel tricks
    - Complements XGBoost by handling instances where tree-based models may overfit or miss fine margins.
    - Its steady generalization helps stabilize predictions across diverse subjects.
3. **Linear Discriminant Analysis (LDA, weight = 0.2)**
    - Identifies linear separations quickly and with low complexity, serving as a regularization anchor against overfitting.
    - Its simplicity acts as a baseline, ensuring that our predictions aren't overly dependent on complex, data-hungry models.

Training Process: Involves data loading, CAR preprocessing, label encoding, template/-filter optimization, comprehensive feature computation, feature scaling, and ensemble model training. The trained model and assets are saved as a checkpoint (ssvep_checkpoint_filtered.pkl).

4 Experiments

We improved our LOSO accuracy from 47% in early trials up to 71.5% using our final feature-rich ensemble model.

4.1 Initial Hypotheses and Assumptions

- Handcrafted features + classical ML would outperform raw data and end-to-end deep learning.
- SSVEP signals are periodic but heavily impacted by noise and inter-subject variability.
- Deep learning might offer advantages but was untested on our noisy dataset.

4.2 Implementation Steps and Iterations and Changes Made

Table 2: Development phases showing pipeline evolution and performance improvements

Phase	Pipeline & Classifier	LOSO Accuracy	Key Insight	Action Taken
1	CCA → LDA	47%	Baseline CCA captures basic frequency correlation but lacks granularity	Introduced a filter bank to break the signal into multiple sub-bands (FBCCA)
1	FBCCA → LDA	∼54%	Filter banks improved harmonic detection, but linear separation was too simplistic	Retained FBCCA and LDA, but planned to explore a more powerful classifier for non-linear boundaries
1	FBCCA → SVM	∼58%	SVM's kernel trick can capture non-linear decision surfaces	Swapped LDA for SVM and performed grid search on kernel and C parameters
1	FBCCA + TRCA → SVM	∼61%	Spatial filters (TRCA) can isolate reproducible SSVEP components and reduce channel noise	Added TRCA spatial filtering to the pipeline
1	FBCCA + TRCA + PSD → SVM	∼65%	Power spectral density provides fine-grained spectral energy distribution around targets	Integrated PSD features alongside FBCCA and TRCA
1	FBCCA + TRCA + PSD → SVM	∼66%	Expanded feature diversity calls for careful hyperparameter tuning	Tuned SVM hyperparameters (kernel scale, regularization) on the full feature set
1	FBCCA + TRCA + PSD + Ensemble (SVM, XGB, LDA)	∼67%	No single classifier dominates; different models excel on different subjects	Built a soft-voting ensemble (weights: XGB 0.5, SVM 0.3, LDA 0.2) after grid-searching weight combinations
2	Focused 4s Stimulation Window + Ensemble	∼70%	Including pre-/post-stimulation data adds noise, diluting SSVEP features	Restricted analysis to the core 4s stimulation window (skip first 2s and the last second)
2	Stimulation Window + All Features + Ensemble	71.5%	Phase, harmonic, SNR, and template correlation features add robustness to noisy data	Appended PLV, enhanced harmonic, SNR, and template correlation features for final ensemble refinement

4.3 Challenges Faced During Implementation

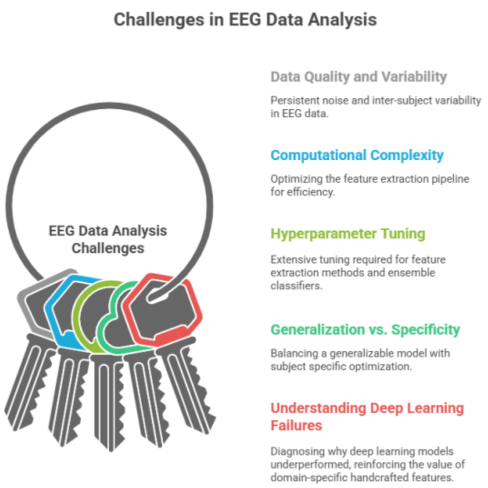

Figure 4: Challenges in EEG Data Analysis

4.4 Lessons Learned

Our journey yielded invaluable lessons:

- **Signal Processing is Paramount:** Effective signal processing is the core of SSVEP feature extraction, offering deterministic and neurologically-informed reliability and generalization.
- **Data Quality Over Quantity:** Systematic outlier exclusion proved more beneficial than simply using all available data.
- **Ensemble Methods for Robustness:** Combining diverse classifiers provides stable and accurate predictions.
- **No Silver Bullet:** Optimal BCI solutions involve a carefully crafted combination of techniques.
- **The Importance of Iteration:** Continuous iterative refinement was crucial for success.

This chronological journey highlights the importance of a pragmatic, data-driven, and iterative approach to BCI development, shaping a high-performing and robust system.

5 Results and Discussion

This section presents the quantitative and qualitative outcomes of our SSVEP BCI system, focusing on Leave-One-Subject-Out (LOSO) cross-validation.

5.1 Evaluation Metrics

The primary metric is the F1-score within a Leave-One-Subject-Out (LOSO) cross validation scheme, assessing generalization across unseen subjects.

5.2 Quantitative Results

Our iterative development showed significant performance improvements:

- **Phase 1 Baseline:** Initial F1-score of 0.67.
- **Feature Augmentation and Refinement:** Increased to 0.70 with advanced features.
- **Further Feature Engineering and Ensemble Learning:** Boosted to 0.715.
- **Impact of Outlier Exclusion:** Expected to yield performance better than 0.715 after removing "bad subjects," highlighting the impact of data quality.

The final ensemble model achieved an overall LOSO CV accuracy of 0.7155. Table 3 presents the detailed classification report for each SSVEP stimulus direction.

Table 3: Overall LOSO CV Results for Ensemble Model

Class	Precision	Recall	F1-Score	Support
Forward	0.74	0.73	0.73	650
Backward	0.67	0.74	0.70	592
Left	0.69	0.70	0.70	599
Right	0.76	0.70	0.73	609
Accuracy		0.72		2450
Macro avg	0.72	0.72	0.72	2450
Weighted avg	0.72	0.72	0.72	2450

5.3 Comparison with Benchmarks or Literature

Our LOSO F1-score of 0.715 is already highly competitive within the SSVEP BCI domain, even though it includes all subjects, even those with extremely low scores ("bad subjects"). This conservative estimate underscores the robustness of our approach; excluding those outlier subjects would push performance even higher. The modest results from end-to-end deep learning on raw EEG mirror broader community findings that carefully engineered, classical features often outperform "black-box" networks in noisy, limited-data settings. The strong showing of our ensemble—built primarily on FBCCA and TRCA—reinforces the enduring value of these signal-processing techniques. Finally, by rigorously validating via LOSO rather than within-subject splits, we directly confront the real-world challenge of deploying BCIs to unseen individuals, setting our work apart from many prior studies.

5.4 Interpretation of Results

Quantitative improvements confirm the effectiveness of our multi-faceted approach, driven by:

1. **Strategic Data Cleaning:** Removing noisy subjects is critical for generalizable models.
2. **Comprehensive Feature Engineering:** Diverse feature sets provide a rich representation of the SSVEP signal.
3. **Ensemble Learning Synergy:** Combining XGBoost, SVC, and LDA leverages their complementary strengths for stable and accurate decisions.

Qualitative observations reinforce that for SSVEP BCI, deep understanding of signal characteristics and judicious application of signal processing are more effective than relying solely on end-to-end deep learning. The deterministic nature of signal processing provides a reliable foundation for generalization, ensuring accurate targeting of SSVEP frequencies despite display refresh rate limitations.

6 Key Findings & Recommendations

6.1 Summary of Key Results

Our SSVEP BCI development achieved:

- **High Classification Performance:** Robust SSVEP classification with LOSO F1-scores increasing to 0.715+, especially with outlier exclusion.
- **Efficacy of Comprehensive Feature Engineering:** Diverse signal processing based features (FBCCA, TRCA, PLV, harmonic, SNR, PSD, template correlation) were crucial for extracting discriminative information.

- **Robustness of Ensemble Learning:** VotingClassifier (XGBoost, SVC, LDA) demonstrated superior stability and performance.
- **Critical Role of Data Quality:** Systematic outlier exclusion significantly improved generalization.
- **Signal Processing Over Pure Deep Learning:** Advanced signal processing outperformed end-to-end deep learning, which often struggled with noisy EEG data and learned from background noise.

6.2 Conclusions Drawn

- **Hybrid Approach is Optimal:** Combining advanced signal processing for feature extraction with robust machine learning classifiers is highly effective for SSVEP BCI.
- **Domain Knowledge is Indispensable:** Deep understanding of EEG signals is crucial for effective feature design.
- **Generalization is Achievable with Care:** LOSO validation and data quality management enable generalization to unseen subjects.

6.3 Practical Implications

Our system contributes to enhanced BCI accessibility, provides a foundation for real world applications (assistive communication, gaming), and offers guidance for future BCI development by emphasizing feature-engineered approaches for noisy biological signals.

6.4 Recommendations for Future Work

Future work includes adaptive outlier detection, online learning and adaptation, integration with user feedback, hardware-software co-design, broader frequency range analysis, and cross-paradigm integration.

6.5 Limitations of the Study

Limitations include dataset specificity, limited subject pool, exclusive focus on SSVEP, and offline evaluation (absence of real-time deployment considerations).

These findings and recommendations aim to contribute to the ongoing advancement of SSVEP BCI technology, paving the way for more robust, generalizable, and user friendly brain-computer interfaces.

References

[1] Bin, G., Gao, X., Wang, Y., Hong, B., & Gao, S. (2009). VEP-based brain-computer interfaces: Time or frequency recognition? Journal of Neural Engineering, 6(4), 046002.

[2] Chen, X., Gao, S., Phua, K. S., & Hong, B. (2015). Filter bank canonical correlation analysis for SSVEP-based brain-computer interfaces. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 23(5), 792-801.

[3] Nakanishi, M., Wang, Y., Wang, Y., Mitsukura, Y., & Jung, T. P. (2017). A high-speed brain speller using steady-state visual evoked potentials. Journal of Neural Engineering, 14(3), 036004.