ELEC5305 Project Proposal

1. Project Title

Audio-Based Event Detection in Soccer Matches Using Deep Learning

2. Student Information

- **Full Name:** Marcellus Ray Gunawan
- **Student ID (SID):** 520038655
- **GitHub Username:** raymrg20
- **GitHub Project Link:** https://github.com/raymrg20/elec5305-project-520038655

3. Project Overview

This project investigates whether audio signals alone can be used to automatically detect key soccer match events such as goals, fouls, and corners. Current event-detection systems rely heavily on video analysis and manual annotation, which are resource-intensive and slow [1]. By leveraging signal processing and deep learning models trained on the SoccerNet V2 dataset, this project aims to design and evaluate an efficient audio-based event detection system [2].

The goal is to compare raw audio feature-based approaches (spectrogram CNN/CRNN) with a baseline Automatic Speech Recognition (ASR) and text-classification approach, demonstrating whether raw signal features can capture critical game events without requiring video or heavy manual input [3].

Comparative analysis of these paradigms will reveal whether critical events in soccer often marked by distinctive crowd reactions, commentator excitement, or sudden audio intensity spikes can be reliably detected from audio streams without visual input or extensive manual labelling. The evaluation will include detection accuracy, computational efficiency, and responsiveness for practical deployment [4].

The outcomes of this project will not only contribute methodological insights but also set the foundation for real-time, lightweight event detection systems that can enhance live broadcast experiences, streamline analytics workflow, and potentially transform how sports are monitored and enjoyed globally. This work addresses the current bias toward video-centric solutions and demonstrates the untapped value in audio signals for sports automation, expanding the scope for scalable and accessible sports analytics.

4. Background and Motivation

Event detection in sports is critical for broadcasting, coaching, refereeing, and fan engagement. Traditionally, annotators manually log key events, but this process is costly and unscalable during tournaments with thousands of hours of footage.

Prior research has focused on video-based detection such as SoccerNet challenge, but video models are computationally expensive [5]. Recently, audio-based methods have emerged as

lightweight alternatives as has been demonstrated using commentary audio transcribed by Whisper ASR and classifying with transformer-based LLMs could detect goals, fouls, and corners with good accuracy [6]. This approach has shown good accuracy in detecting key soccer events like goals, fouls, and corners, leveraging the rich contextual cues present in commentary and crowd reactions.

However, that work treated the problem primarily as text classification [6]. This project proposes to extend the idea by returning to the signal domain using raw audio spectrograms and learned features to test whether whistles, crowd voices, and acoustic patterns provide a strong basis for event detection. This makes the project more aligned with audio signal processing and deep learning, bridging theory with practical sports applications.

By grounding the project in audio signal processing and deep learning, it bridges the foundational theory of audio analysis with practical sports applications. This fusion promises enhanced robustness, lower latency, and scalability for real-time event detection systems, positioning audio-based methods as a valuable complement or alternative to video-centric pipelines.

5. Proposed Methodology

a. Dataset:

- Use SoccerNet V2 (500 matches, 764 hours, ~300,000 labeled events).
- Extract the broadcast audio tracks aligned with annotated events (Goals, Fouls, Corners).

b. Preprocessing & Feature Extraction:

- Segment audio into 15 s and 30 s windows around events.
- Compute Short-Time Fourier Transform (STFT) and log-mel spectrograms.
- Augment data with noise injection, pitch/time-shifting to handle variability.

c. Modeling Approaches:

- Signal-based deep learning: CNN/CRNN on spectrograms.
- Baseline (text-based): Whisper ASR → commentary transcription → transformer classifier.

d. Evaluation Metrics:

- Event-level Precision, Recall, F1-score.
- Confusion matrices to study misclassifications.
- Computational efficiency (training/inference time).

e. Tools & Platforms:

- Matlab (Deep Learning Package) or Python (PyTorch) for modeling.
- Google Colab / HPC cluster for training.
- GitHub for code/documentation, GitHub Pages for project site.

6. Expected Outcomes

- Working prototype: A deep learning model that can classify audio clips into event categories (Goal, Foul, Corner).
- Baseline comparison: Evaluate ASR+LLM vs. spectrogram CNN/CRNN approaches.
- Performance metrics: Precision, Recall, F1-score.
- Deliverables:
    - Open-source codebase (GitHub).
    - Technical report and GitHub Pages site.
    - Short demo video of the system in action.

7. Timeline (Week 6 – 13)

Week	Tasks
6-7	Literature Review and Extracting the Dataset from SoccerNet
8-9	Preprocessing and Implement CNN/CRNN on spectrograms of the audio signals
10-11	Optimize models (hyperparameter tuning, augmentation experiments)
12	Evaluate using confusion matrix and error analysis
13	Finalise report, code, and video demo

References

[1] J. Bischofberger, A. Baca, and E. Schikuta, "Event detection in football: Improving the reliability of match analysis," *PloS one*, vol. 19, no. 4, pp. e0298107–e0298107, Apr. 2024, doi: https://doi.org/10.1371/journal.pone.0298107.

[2] S. Giancola, "SoccerNet-v2," *Github.io*, 2021. https://silviogiancola.github.io/SoccerNetv2/

[3] H. Ilgaz, B. Akkoyun, Ö. Alpay, and M. A. Akcayol, "CNN Based Automatic Speech Recognition: A Comparative Study," *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, vol. 13, p. e29191, Aug. 2024, doi: https://doi.org/10.14201/adcaij.29191.

[4] S. Gautam, C. Midoglu, S. Shafiee Sabet, D. B. Kshatri, and P. Halvorsen, "Soccer Game Summarization using Audio Commentary, Metadata, and Captions," *Proceedings of the 1st Workshop on User-centric Narrative Summarization of Long Videos*, Oct. 2022, doi: https://doi.org/10.1145/3552463.3557019.

[5] S. Challenges, "SoccerNet - Challenges," *Soccer-net.org*, 2025. https://www.soccer-net.org/challenges.

[6] J. Teklemariam, "Automatic Detection of Soccer Events using Game Audio and Large Language Models." Accessed: Sep. 07, 2025. [Online]. Available: https://home.simula.no/~paalh/students/2024-NMBU-JoelYacobTeklemariam.pdf