SoundSafe.ai Attribution Layer

1. Introduction

This document describes the SoundSafe attribution layer, which is the system that ties together various components to provide a robust and comprehensive solution for embedding, tracking, reporting, and attributing audio content across various platforms. This layer orchestrates data flow, processing logic, and user interactions from metadata management to real-time analytics and alerting, designed to ensure efficient, scalable, and secure operations within the SoundSafe framework.

2. Core Components of the Attribution Layer

The attribution layer is not a single component but a synthesis of several interconnected modules that work together harmoniously:

- **Metadata Handling Module:**
  - **Purpose:** To collect, validate, clean, and incorporate metadata from various sources such as direct imports from Spotify and YouTube Music via an API, databases, and manual user inputs.
  - **Functions:** Includes job queueing for batch operations, a robust error handling and reporting mechanism, handling partial successes, provides various pre-processing functions, progress tracking and customizable templates.
  - **Optimization**: Uses best practices in metadata structure to ensure scalability, maintainability, security and interoperability across various systems by implementing a hierarchical system for efficient access. The data management structure also is designed to track lineage for transparency and ease of auditing.

- **Parallel Hybrid Model (PHM):**
  - **Purpose:** A novel architecture designed to combine the embedding of robust and imperceptible watermarks into audio, and the extraction of unique acoustic fingerprints of audio content for real-time detection, verification and tracking.
  - **Structure:** The core innovation of PHM lies in it's a parallel approach to analysis and watermark embedding where:
    - *Perceptual Branch:* Uses a light-weight CNN to extract critical perceptual elements from an audio to minimize performance overhead and computational cost. This ensures that high fidelity is retained in the final watermarked audio while also ensuring lower resource overhead when running across different edge locations and platforms.
    - *Technical Branch:* Employs optimized RNN structures for robust watermark embedding that focus on encoding data with minimum bit error rate. This separation ensures the model is flexible for multiple use cases.
    - *Fusion Layer:* Integrates both embeddings to provide a holistic and robust model for diverse use cases. Using self attention models ensures optimized weighting for watermark representation.
  - **Integration** The PHM is integrated in multiple ways to ensure fast detection (via light version) as well as comprehensive encoding/decoding mechanisms (for batch operations). PHM operates in both client side and server side implementations for edge use cases, which ensures optimal usage of resources depending on the type of use-case and device limitations. This approach further enables for incremental enhancements to each branch individually, and better flexibility when optimizing performance.
  - **Optimization**: PHM is continuously improved by data collected in each tracking, detection and embedding mechanism to fine-tune itself. The model utilizes distributed and federated learning techniques where possible to be able to learn based on regional bias that may exist within the system for enhanced results.

- **Scanning and Detection System:**
    - **Purpose:** To monitor various platforms in near-realtime by detecting watermarked content and match unwatermarked audio against a DB. Uses AI for both improved scanning results and extraction.
    - **Mechanism:** Utilizes the PHM to process audio across different platforms using simplified versions where needed to match various computational and hardware resource limitations. Fingerprints are used to compare new audios with known audios, with results going to a comprehensive usage reporting tool. The system intelligently distributes scans based on need, prioritizes the most relevant audio, and intelligently assigns compute and resources depending on device or hardware capability
    - **Optimization**: Uses AI for more adaptive thresholding, self-learning fingerprint detection, and data cleaning. Includes multi-location distribution using geographically close data centers and microservices for scalability and to minimize network costs.

- **Tracking and Reporting Engine:**
    - **Purpose:** To provide granular tracking and aggregated usage data with robust reporting features. Generates customized real-time notifications.
    - **Capabilities:** Offers user customized real time tracking, data summaries, real time visualization (such as dashboards). Also can generate summarized periodic reports, including trend analysis using AI powered engines.
    - **Optimization**: Streamlines data with various formats for easier access by different external systems, with robust role based user authentication management, while providing comprehensive tools for audit and compliance. Data can be filtered, sorted and accessed via a specific metadata field to further improve the workflow of end users. Supports batch export and analysis of very large datasets as well.

- **Payload Encoding:**
    - **Purpose**: To handle dynamic and variable payloads related to specific metadata such as DDEX standards for the music industry, so systems are easily interoperable across the whole music industry ecosystem. The system prioritizes using hierarchical systems for better encoding and faster performance while also supporting backward compatibility, flexibility and scaling as the data payload increases in the future.
    - **Mechanism**: Uses multiple optimization algorithms to reduce overhead of bits. Dynamically allocating variable length encoding where applicable using strategies such as LZ compression, Huffman encoding and custom lossless coding and indexes. All essential parts of payload for watermarking will have a very low bit error rate while non-essential parts can have much smaller fidelity depending on the requirement for different parts.
    - **Optimization:** By breaking metadata into core fields and non core fields enables much higher robustness, and flexibility when embedding large datasets, and hierarchical encoding allows the system to allocate the right amount of space and prioritize important parts of metadata during the encoding/decoding phase. This also prioritizes fast watermark embedding/detection over retrieval of all metadata, with metadata available for reports or compliance related analysis when needed in the background

- **Data Layer**
    - **Purpose:** This layer acts as a common ground to transfer data across diverse use-cases and various parts of the infrastructure in an efficient, reliable, fast and scalable manner. Also facilitates storing data in optimal formats for different applications.
    - **Mechanism:** Employs various types of scalable databases (both traditional and no-SQL type) along with a robust and fault-tolerant architecture so that services can access all information without data integrity and performance problems. Uses a caching mechanism with CDN to ensure consistent information across the world. Provides fault tolerance and guarantees data consistency as various servers go down or have performance issues. Has distributed data queues to ensure consistent message and data transfer for inter-module communication. Uses geo location and access time for local data prioritization, so systems can intelligently adapt based on specific geo locations. Uses encryption to make sure only authorized users/applications can access various systems.
      * **Optimization**: Is fully modularized with APIs for extensibility and data type transformations are handled in modular fashion via a system of well-defined interfaces for easy adaptation to diverse use cases and data transformations and security needs of new applications and integrations with third party tools.

3. Interaction of Components: The Attribution Process

The attribution layer works through the following steps:

1. **Metadata Ingestion:** Relevant information about a track (e.g., artist, title, ISRC) is ingested into the metadata handling module.
2. **Watermark Embedding:** Using the PHM the system will encode watermark data (based on ingested data) onto the audio file. This generates the watermarked audio.
3. **Distribution:** Watermarked audio is distributed to platforms, streaming services, etc.
4. **Real-time Monitoring:** The scanning and detection modules monitor platforms and channels (Youtube, Spotify and many others) for any traces of previously watermarked audio. The simplified PHM is utilized to quickly perform this operation in the client side to save on CPU, bandwidth and computational load.
5. **Fingerprint generation** Every detected audio, either watermarked or not, has it's fingerprinted based on different use-cases. Watermarked audio detection for compliance, and tracking, while unwatermarked data is collected and compared for metadata verification. Data is stored in the data management layer in multiple geographic locations for redundancy, low latency and better performance.
6. **Matching and Identification:** Scanned audio is processed and compared with records in data management, and then matched using smart thresholding that adjust for differences in various audio, device or environment characteristics, using cosine or Euclidean similarity metrics. The PHM architecture combined with the efficient matching mechanism makes the detection results more reliable.
7. **Usage Tracking:** The system records detailed data when watermarked audio is detected. Granular data, and location data for compliance and reporting are collected. This includes usage time stamps, number of users, geographic location and the nature of the usage (Full use, partial or micro use).
8. **Metadata Reconciliation**: By performing fingerprint detection on all audios, the data layer is able to match non-watermarked data against known metadata with their fingerprints from previously processed datasets, in addition to metadata gathered from various sources. This ability to automatically match can reduce time spent identifying content, or resolving ownership or compliance issues.
9. **Hierarchical data transfer, analysis, and reporting**: The data, usage data and results of analysis, from all detection points across the world are transferred to central database systems using geo location based prioritization. Each server prioritizes handling data closer to where the source of the data exists, and transfers are done asynchronously, to not overload a single area, by summarizing the data before transferring for more efficiency. This results in near real time data aggregation, without any data integrity concerns, and lower cost and better overall performance across all geographical areas.
10. **Reporting and Analytics:** The system aggregates collected usage data and creates reports and visualizations with key insights including alerts based on user defined criteria.

4. Key Attributes of the Attribution Layer

- **Scalability:** Designed for global deployment and millions of transactions with a distributed processing framework.
- **Real-Time:** Provides actionable usage data using distributed processing and asynchronous queues to ensure consistent data transfers across many geographical regions.
- **Robustness:** PHM, multi-point error handling with self learning, fault tolerance.
- **Efficiency:** Optimized algorithms and data structures, and a multi-tiered approach of having light weight systems and full heavy processing based on device capacity to minimize cost and operational overheads.
- **Interoperability:** Support for industry standards ensures compliance, modular API architecture enables integration of external tools and systems.
- **AI Powered Metadata and Matching:** Provides a scalable, self-learning mechanism to improve metadata accuracy, fingerprint identification and watermarking that reduces false positive and false negative reporting by taking into account biases introduced by data.
- **Dynamic Bit Management System:** A robust hierarchical encoding and data handling system allows flexible management of metadata for different payloads.

The SoundSafe attribution layer is not a single component but a collection of systems orchestrated to provide a unified and highly effective solution that ties everything together. It provides robust embedding, real time global tracking and reporting with a focus on scalability, security, interoperability and fault-tolerance. This modular design allows for continuous innovation with new technologies to enable seamless interoperability with diverse types of audio platforms across a global environment.