Video Notes

Overview of Shazam's Audio Recognition Technology

Introduction to Audio Fingerprinting

- **Concept**: Audio fingerprinting is akin to creating a unique DNA profile for every song.
- **Purpose**: Allows Shazam to identify songs from brief audio snippets.

Process of Song Identification

1. **Audio to Spectrogram Conversion**
    - **Definition**: A spectrogram is a visual representation of the frequency spectrum of sound as it varies with time.
    - **Method**: Audio is converted into a spectrogram by capturing frequency content over time.
    - **Details**:
        - Peaks in the spectrogram serve as the song's unique identifiers.
        - Audio is split into overlapping segments using a sliding window technique to analyze smaller frequency components.
2. **Peak Extraction and Filtering**
    - **Technique**: Application of a filter with six logarithmic frequency bands, designed to mimic human ear sensitivity.
    - **Frequency Sensitivity**:
        - Human ears are less sensitive to frequencies below 500 Hz.
        - More attuned to mid to high frequencies (500 to 2000 Hz).
    - **Dynamic Thresholding**:
        - Calculation of the average magnitude of the six strongest frequencies to set a dynamic threshold.
        - Frequencies below this threshold are discarded to ensure only significant peaks are retained.
3. **Time Coherence Analysis**
    - **Purpose**: To evaluate how well the time stamps of audio snippet peaks align with those of each candidate song.
    - **Outcome**: The song with the highest time coherence is identified as the match.

Implementation of Technology

1. **Front-End Development**
    - **Choice of Technology**: Opted for React over vanilla JavaScript for ease after complex backend development.
    - **User Interaction**:
        - Users record a short audio clip via microphone or computer.
        - Audio snippet is encoded in Base64 format and sent to the server.
2. **Server-Side Processing**
    - **WebSocket Usage**: For real-time communication between the front end and server.
    - **Matching Process**:
        - Extracted audio fingerprints are compared against a database of song fingerprints.
        - Matches are determined based on time coherence and the consistency of peak timestamps within a 100-millisecond tolerance.
3. **Data Handling**

- **Metadata Storage**: Each song match is stored with metadata, earliest anchor time, and a coherence score.

Conclusion

Shazam's audio recognition technology leverages sophisticated digital signal processing techniques including spectrogram analysis, dynamic thresholding, and time coherence analysis to accurately identify songs from audio snippets. This process not only showcases advanced technical implementation but also a user-friendly interface facilitated by modern web technologies.