A System and Method for Automated, Lyric-based Image Generation as Synchronized Song Accompaniment

BACKGROUND OF THE INVENTION

[0001] Image-Based Musical Accompaniment: The history of creating a visual aspect to accompany music is long and varied, ranging from ancient Greek drama to the advent of music television channels like MTV. Today, a popular song that includes lyrics is often accompanied by one or more music videos, which present images or video synchronized and associated with the song's lyrics. Listeners find entertainment value in seeing images associated with the music they are listening to.

[0002] Text-Based Generative AI: Text-based generative AI systems have received increasing attention in the last decade, as their abilities have improved to meet a wider range of requirements. Such Large Language Models (LLMs) are typically provided with a "system prompt", which guides their behavior in creating a text-based response to a "user prompt", along with other information. Because the system prompt can be flexibly defined, these LLMs are capable of a wide range of useful behaviors, including reformatting, modifying, and interpreting the user query, as well as creating novel textual content.

[0003] Image-Based Generative AI: Recent advancements in image-based generative AI have enabled the creation of visually-appealing content from textual descriptions. These systems are models trained to generate images that are thematically consistent with the provided text, enabling a new dimension of multimedia content.

SUMMARY OF THE INVENTION

[0004] This invention provides a method for creating an image-based musical accompaniment to a lyric-containing song. Via a text-based LLM, each subsequent lyric is converted into an appropriate prompt for an image-based generative AI system. A variety of techniques ensure that the created images are thematically associated with their corresponding lyrics. After each image is generated, the created images are displayed as a visual accompaniment to the audio component of the music, allowing the user to experience the newly-formed visual accompaniment.

BRIEF DESCRIPTION OF DRAWINGS

[0005] FIG. 1 – Frontend Mockup: This diagram depicts the user interface for displaying synchronized lyrics and images. It includes relevant frontend components, such as an Image Display Area (110), User Input Module (100), and Audio Playback Controls (140).

[0006] FIG 2. – System Architecture: This diagram depicts the structural design of the system, illustrating how modules such as Frontend (Web Page) (210), Backend (Server) (220), and Lyric & Metadata Fetching (230) interact with each other.

[0007] FIG 3. – System Flow: This diagram depicts the step-by-step process of how the system processes a song, generates a global theme, and creates and presents synchronized images. Alongside the system flow column, another column, 'Example Lyric Flow', illustrates the process with an example song for increased clarity.

DETAILED DESCRIPTION OF THE INVENTION

[0008] This invention pertains to the field of multimedia content generation and synchronization, with a particular focus on the automated creation of image-based accompaniments for musical works. Leveraging advancements in text-based generative AI and image-based generative AI, this invention enhances the user experience of listening to music by providing synchronized visual content that has appropriate semantic relation to the metadata and lyrics of a musical work.

[0009] The primary objective of this invention is to create a cohesive and immersive multimedia experience where dynamically generated images align with the lyrical content of a song, synchronized with the song's audio and lyrics.

[0010] The invention is implemented through an integrated system comprising a user-facing interface and a backend processing framework. These components collaboratively manage user inputs, retrieve relevant musical metadata and lyrics, analyze the lyrical content to generate thematic elements, create image generation prompts, produce synchronized visual content, and deliver a cohesive audiovisual experience. The system consists of six key components, each contributing to synchronized multimedia content generation.

[0011] The first component is the User Interface Module (100), as shown in FIG. 1, and represented in User Input (211), shown in FIG. 2. This module facilitates user input of identifiers or links corresponding to specific musical works. This module captures and processes user inputs to initiate the multimedia generation process.

[0012] The second component is the Data Acquisition Module (230), as shown in FIG. 2. This module accesses and retrieves pertinent metadata and time-synced lyrics associated with the selected musical work from various data sources.

[0013] The third component is the Global Theme Generation Module, as shown in (241) and in FIG. 2. This module utilizes text-based artificial intelligence to analyze retrieved metadata (complete lyrics, genre, album cover, artist name, album name, etc.) and extract thematic elements. These thematic elements form a global thematic framework (interchangeably referred to as “Global Theme” throughout), which guides the generation of image prompts, ensuring synergy and cohesion amongst generated images.

[0014] The fourth component is the Prompt Generation Module (250). This module generates subsequent image generation prompts, utilizing both the song-wide Global Theme (241), as well as the individual lyric in question, to generate cohesive image generation prompts for each time-stamped lyric (243). Due to the use of the global theme, these prompts typically describe relatively similar color palettes, art styles, themes, and motifs.

[0015] The fifth component is the Image Generation Module (244). This module utilizes image generation algorithms, in one implementation Stable Diffusion, to produce visual representations from the generated prompts. This module processes each derived prompt to create images that are thematically consistent with the song’s narrative and emotional tone.

[0016] The sixth component is the Multimedia Synchronization Module. This module coordinates the timing of image displays with the audio playback of the musical work. This module ensures that each generated image aligns accurately with its corresponding lyrical timestamp. Additionally, it manages the display of scrolling lyrics (120), FIG. 1., alongside the audio content (130), and visual content (110), enhancing the immersive audiovisual experience for the user.

[0017] By seamlessly integrating these components, the system automates the creation of synchronized visual accompaniments to musical works. Users can experience their selected songs enhanced with dynamically generated images and synchronized scrolling lyrics that reflect and complement the lyrical and auditory elements, offering a novel and engaging way to interact with music.

[0018] The invention can be implemented using various programming languages and frameworks to achieve the desired functionality. In one embodiment, the system comprises a frontend developed using React.js and a backend implemented with Python using Flask and

Celery. This architecture allows for asynchronous processing and efficient handling of user requests.

[0019] As shown in FIG. 3, the system flow involves several steps to synchronize the multimedia content. The user enters a song identifier, such as "House of the Rising Sun," and the system validates the input and retrieves the corresponding metadata and lyrics. The backend initiates a processing task to handle the AI computations. The AI model analyzes the full lyrics and metadata to generate a global theme. For example, it may determine an art style of "Victorian mansion in New Orleans, foggy night, wrought iron gates, overgrown garden, hauntingly beautiful, detailed hand-drawn style" For each lyric, the system generates an image prompt, such as "Victorian mansion in New Orleans, foggy night, wrought iron gates, overgrown garden" for the lyric "There is a house in New Orleans." The image generation model creates an image based on the prompt, and the generated image is saved along with its file path and timestamp.

[0020] The system prepares the data for playback, and during playback, the frontend retrieves the necessary data from the backend to display the images and lyrics in synchronization with the audio playback. The images and lyrics are displayed in sync with the audio, providing an immersive experience.

[0021] The described system offers several benefits. By leveraging AI models, the system automates the creation of visual accompaniments, reducing the need for manual effort. Users can generate unique visual experiences for any song with lyrics, enhancing their engagement with the music. The use of a global theme ensures that all images align aesthetically, providing a cohesive visual narrative. The modular design allows for easy integration of new AI models and adaptation to different platforms or languages.

[0022] Alternative embodiments are possible. While the embodiment described utilizes specific technologies and frameworks, the system can be implemented using other programming languages such as JavaScript, Ruby, or Java, depending on the requirements. Various text-based and image-based AI models can be integrated, including proprietary or open-source solutions. The system can also be adapted to work with other forms of media, such as podcasts or audiobooks, where synchronized visual content enhances the user experience.

[0023] The system can be deployed on various hardware configurations, ranging from local servers to cloud-based infrastructures. Utilizing cloud services allows for scalable processing power to handle multiple users and high-demand periods. It is essential to implement security

measures to protect user data and comply with privacy regulations, such as secure data transmission, user authentication, and data storage compliance.

Claims

1. A system for generating synchronized multimedia content, comprising:
    a. A lyrics retrieval module configured to obtain time-synced lyrics associated with a selected musical track, with optional metadata retrieval for additional information such as genre, album cover, artist name, and album name;
    b. A thematic analysis module configured to analyze the retrieved lyrics to extract thematic elements;
    c. An image prompt generation module configured to generate image prompts based on both the lyrical content as well as input from the extracted thematic elements;
    d. An AI-driven image generation module configured to create visual content based on the generated image prompts;
    e. A synchronization module configured to align the generated visual content with the audio playback of the musical track, thereby creating an immersive audiovisual experience.

Abstract

The disclosed invention describes a method and system for creating an image-based accompaniment to any work of music synchronized to specific lyrics. This method utilizes a text-based generative Artificial Intelligence (AI) system to iteratively convert subsequent lyrical components into prompts for a separate image-based generative AI system, guided by a **global thematic framework**. The **global thematic framework** ensures that the generated images are thematically aligned with the given lyrics, producing corresponding visuals that enhance the auditory experience. These images are then presented as a visual accompaniment alongside the audio component of the music, providing an immersive and contextually rich multimedia experience.

Drawings

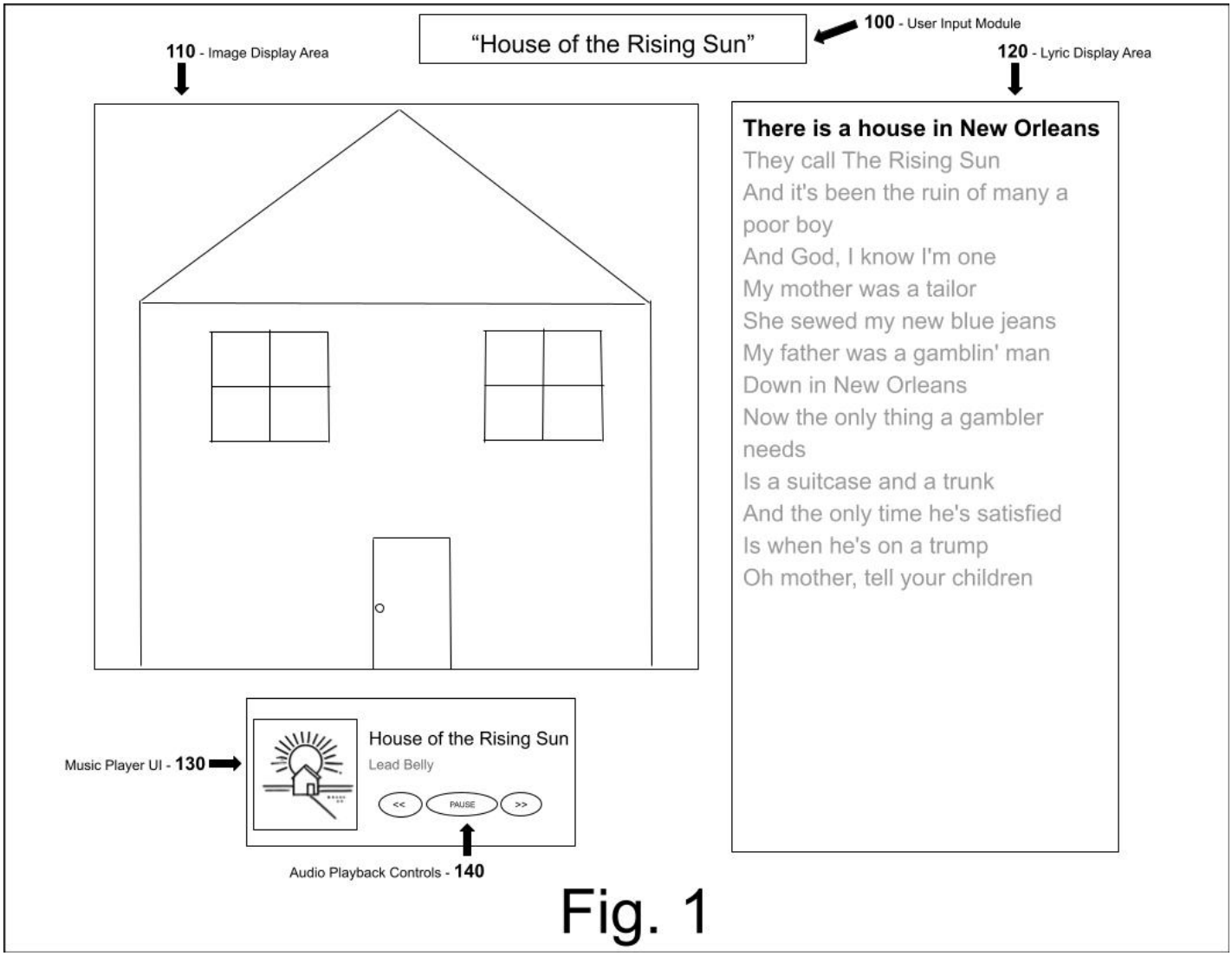

Fig. 1

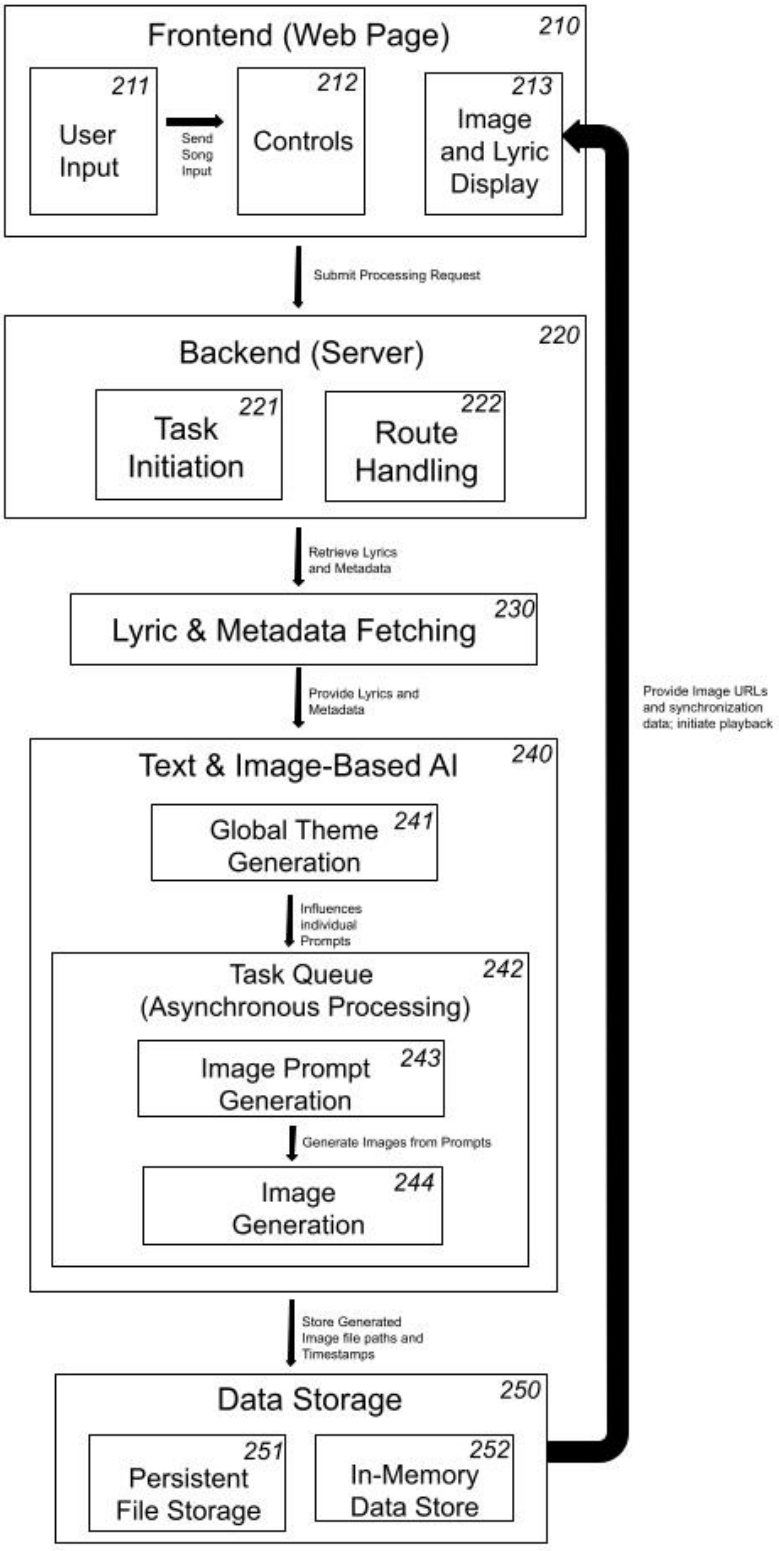
Frontend (Web Page)
210
211
User Input
Send Song Input
212
Controls
213
Image and Lyric Display
Submit Processing Request
Backend (Server)
220
221
Task Initiation
222
Route Handling
Retrieve Lyrics and Metadata
Lyric & Metadata Fetching
230
Provide Lyrics and Metadata
Text & Image-Based AI
240
Global Theme Generation
241
Influences individual Prompts
Task Queue (Asynchronous Processing)
242
Image Prompt Generation
243
Generate Images from Prompts
Image Generation
244
Store Generated Image file paths and Timestamps
Data Storage
250
Persistent File Storage
251
In-Memory Data Store
252
Provide Image URLs and synchronization data; initiate playback

Fig. 2

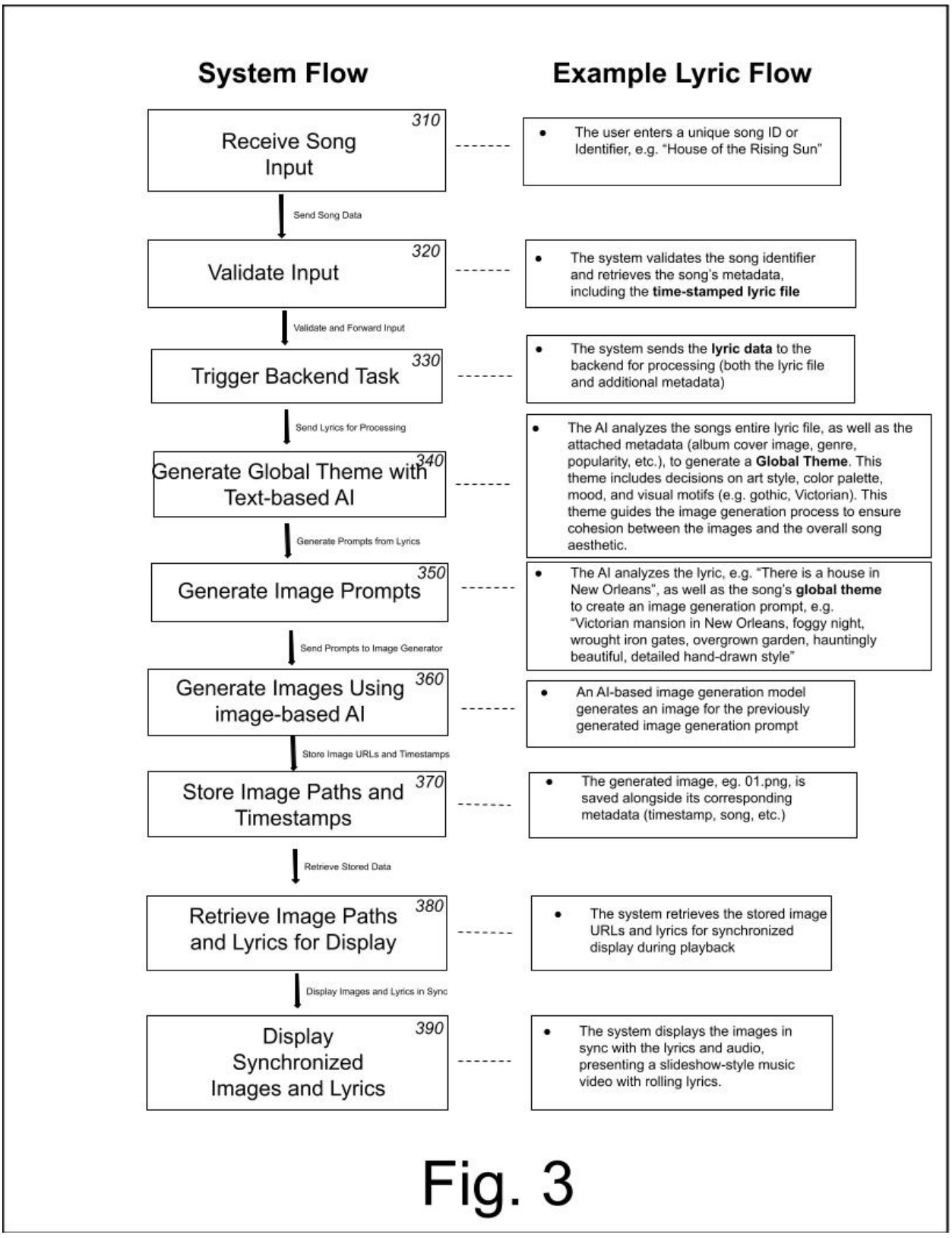
System Flow
Example Lyric Flow
310
Receive Song Input
The user enters a unique song ID or Identifier, e.g. "House of the Rising Sun"
Send Song Data
320
Validate Input
The system validates the song identifier and retrieves the song's metadata, including the time-stamped lyric file
Validate and Forward Input
330
Trigger Backend Task
The system sends the lyric data to the backend for processing (both the lyric file and additional metadata)
Send Lyrics for Processing
340
Generate Global Theme with Text-based AI
The AI analyzes the songs entire lyric file, as well as the attached metadata (album cover image, genre, popularity, etc.), to generate a Global Theme. This theme includes decisions on art style, color palette, mood, and visual motifs (e.g. gothic, Victorian). This theme guides the image generation process to ensure cohesion between the images and the overall song aesthetic.
Generate Prompts from Lyrics
350
Generate Image Prompts
The AI analyzes the lyric, e.g. "There is a house in New Orleans", as well as the song's global theme to create an image generation prompt, e.g. "Victorian mansion in New Orleans, foggy night, wrought iron gates, overgrown garden, hauntingly beautiful, detailed hand-drawn style"
Send Prompts to Image Generator
360
Generate Images Using image-based AI
An AI-based image generation model generates an image for the previously generated image generation prompt
Store Image URLs and Timestamps
370
Store Image Paths and Timestamps
The generated image, eg. 01.png, is saved alongside its corresponding metadata (timestamp, song, etc.)
Retrieve Stored Data
380
Retrieve Image Paths and Lyrics for Display
The system retrieves the stored image URLs and lyrics for synchronized display during playback
Display Images and Lyrics in Sync
390
Display Synchronized Images and Lyrics
The system displays the images in sync with the lyrics and audio, presenting a slideshow-style music video with rolling lyrics.

Fig. 3